

N° d'ordre:

N° de série:



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET

DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ECHAHID HAMMA LAKHDAR D'EL-OUED

FACULTÉ DES SCIENCES EXACTES

DEPARTEMENT D'INFORMATIQUE

MEMOIRE DE FIN D'ETUDE

MASTER ACADEMIQUE

Domaine: Mathématiques et Informatique

Filière: Informatique

Spécialité: Systèmes Distribués et Intelligence Artificiel (SDIA)

THEME

Analyse de la texture des images mammaires par une fusion des lois de Zipf et des matrices de co-occurrence des niveaux de gris dans un processus de classification des tumeurs mammaires.

Présenté par:

KAHLAOUI Asma

MASMOUDI Souheila

Soutenu le XX Septembre 2019

Devant le jury composé de:

Dr Hamoud Meriem MC (B)

M..... MAA l'Université d'El Oued

M..... MAA l'Université d'El Oued

Encadreuse

Président

Examineur

Année universitaire: 2018/2019



REMERCIEMENT

*Tout d'abord, Nous tenons à remercier ALLAH pour nous avoir donné le courage,
la volonté pour réussir et réaliser ce modeste travail.*

*Nous tenons à exprimer toute notre gratitude à notre encadreur Melle D.
HAMOUD MERIEM, pour sa patience, sa disponibilité et
surtout ses judicieux conseils qui ont contribué à alimenter notre réflexion, ainsi que
ses encouragements lors de la réalisation de notre mémoire.*

*Nous adressons nos sincères remerciements à Mrs les membres de jury qui nous
honore par leurs présences en qualité*

*Nous remercions infiniment tous les enseignants qui nous ont aidés durant tout notre
cycle d'études.*

*Un grand merci à tous ceux qui ont contribué de près ou de loin pour l'élaboration de
ce travail.*

Nous remercions tous nos collègues et amis sans exception

*Nous remercions tous ceux qui nous ont aidés de près ou de loin pour élaborer ce
travail.*

Merci au lecteur intéressé par notre travail.

Dédicace

À ma mère, à cause de sa patience et de ses encouragements lors de dures épreuves et de prières qui m'apportent bonheur et succès.

À mon père, pour son grand amour, ses encouragements, son sens du devoir et son sacrifice pour la réussite de mes études.

À mon époux qui ne manque pas à me donne de courages sans arrêt.

À mon bébé qui m'envoyé l'espoir à tous les moments.

À mes sœurs, pour mes frères, qui sont toujours à mes côtés, prêts à m'aider.

À mes chères amies, à tous ceux qui m'aiment et à tous ceux que j'aime.

Pour chaque personne que je connais de près ou de loin

Kahlaoui Asma



Dédicace

A mon mari

A mes enfants

A mes sœurs

A mes frères

A mes amies

A mes collègues

Masmoudi Souheila



RÉSUMÉ

Vu la difficulté de la modélisation des images, possédant une structure complexe, au biais de simples relations linéaires; ceci a renforcé l'utilisation de la non-linéarité. En effet, notre problématique de recherche consiste à appliquer les lois puissance: Zipf et Zipf inverse à l'analyse des images mammaires. D'une part, les lois de Zipf caractérisent la complexité structurelle de la texture d'image par la modélisation de la répartition statistique de la fréquence d'apparition des motifs selon une distribution en lois puissance. Par ailleurs, nous fusionnons les descripteurs obtenus avec ceux générés suite à l'analyse des images mammaires par la matrice de co-occurrence d'Haralick. De plus, une phase de sélection des descripteurs les plus discriminants est réalisée par les algorithmes génétiques. Dans ce sens, nous avons conçu un système d'indexation et de recherche des mammographies par le contenu (CBMIIR) qui renforce la performance du diagnostic assisté par ordinateur au niveau de l'étape de la présentation du diagnostic aux radiologues. En effet, ces derniers sont plus confiants d'un diagnostic basé sur des cas, diagnostiqués préalablement, similaires au cas en cours d'analyse plutôt que le résultat abstraits généré par un classifieur. L'évaluation de l'approche proposée basée la technique de classification Template-matching a donné des performances encourageantes.

Mots clés : Analyse d'image, Vision par ordinateur, Loi de Zipf, Loi de Zipf inverse, classification, Indexation, Recherche par le contenu, Matrice de co-occurrence, Algorithmes génétiques, Template-matching.

Abstract

Given the difficulty of modeling images, possessing a complex structure, through simple linear relations; this reinforced the use of non-linearity. Indeed, our research problem is to apply the power laws: Zipf and inverse Zipf to mammogram images analysis. Firstly, Zipf's laws characterize the structural complexity of the image texture by modeling the statistical distribution of patterns frequency of appearance as power law distribution. Otherwise, we have performed a fusion of the obtained descriptors with those generated once analyzing mammogram images with the Haralick co-occurrence matrix. In addition, a selection phase of the most discriminating descriptors is carried out by genetic algorithms. Furthermore, we have proposed a Content-based mammogram image indexing and retrieval system (CBMIIR) that boosts the performance of a computer aided diagnosis (CADx) at the stage of providing the diagnostic to radiologists, indeed, radiologists feel more confident in their diagnosis decision based upon case-adaptive classification where similar known cases, to the one under analysis, are retrieved and displayed from indexed databases; rather than the abstract result generated by a classifier. The evaluation of the proposed approach based on the Template-matching classification technique gave encouraging results.

Keywords: Image analysis, Computer vision, Zipf's law, inverse Zipf's law, Classification, Indexing, Retrieval by content, Co-occurrence matrix, Genetic algorithms, Template-matching.

المخلص

نظرا لصعوبة نمذجة الصور ، ووجود بنية معقدة ، عن طريق العلاقات الخطية ؛ هذا يعزز استخدام غير الخطية. في الواقع ، تتمثل مشكلة بحثنا في تطبيق قوانين القوة: زيب أف وزيب أف معكوس في تحليل صور الثدي. من ناحية ، تميز قوانين زيب أف التعقيد الهيكلي لقوام الصورة من خلال نمذجة التوزيع الإحصائي لتكرار حدوث الأنماط وفقاً للتوزيع في قوانين الطاقة من جهة أخرى ، قمنا بدمج واصفات تم الحصول عليها مع تلك التي تم إنشاؤها بعد تحليل صور الثدي بواسطة مصفوفة التعاون أغاليك بالإضافة إلى ذلك ، يتم تنفيذ مرحلة اختيار أكثر الواصفات التمييزية بواسطة الخوارزميات الجينية. وبهذا المعنى ، قمنا بتصميم نظام فهرسة وتصوير الثدي بالأشعة السينية القائم على المحتوى (CBMIR) الذي يعزز أداء التشخيصات بمساعدة الكمبيوتر في مرحلة تقديم التشخيص إلى أخصائي الأشعة. في الواقع ، هم أكثر ثقة في حكم التشخيص بناءً على الحالات التي تم تشخيصها سابقاً على غرار الحالة التي يتم تحليلها بدلاً من النتيجة المجردة الناتجة عن المصنف. أعطى تقييم النهج المقترح المبني على تقنية التصنيف بمطابقة القوالب نتائج مشجعة.

كلمات البحث: تحليل الصور؛ رؤية الكمبيوتر، قانون زيب أف ؛ قانون زيب أف معكوس ؛ التصنيف ، الفهرسة ، البحث حسب المحتوى ، مصفوفة التعاون أغاليك ، خوارزميات الجينية ، مطابقة القالب.

LISTE DES MATIÈRES

Résumé

Abstract

ملخص

Liste des matières

Liste des figures

Liste des tableaux

Liste des acronymes

Introduction générale.....1

Chapitre I Analyse d'image et vision par ordinateur

I.1 Introduction.....3

I.2. Concept de l'analyse d'image4

I.2.1 Les différentes phases de l'analyse d'une image.....4

I.2.1.a Analyse de bas niveau d'image.....5

I.2.1.b Analyse d'intermédiaire niveau d'image.....7

I.2.1.c Analyse de haut niveau d'image.....7

I.3 Concept de la vision par ordinateur.....7

I.4 Analyse de la texture d'image.....8

I.4.1 Définition de la texture.....8

I.4.2 Caractérisation de la texture d'image9

I.4.3 Domaines d'application de l'analyse de la texture.....11

I.4.3 Conclusion.....13

Chapitre II : les lois puissance

II.1. Introduction.....	14
II.2. Qu'est-ce qu'une loi puissance?.....	15
II.3. Les lois puissance existantes	15
II.3.1 La loi de Pareto.....	15
II.3.2. Loi de Benoit Mandelbrot.....	16
II.3.3. La loi de Robert GIBRAT.....	17
II.3.4. La loi de Zipf	17
II.3.5 La loi de Zipf inverse	18
II.4 Application des lois de Zipf et de Zipf inverse en analyse d'image.....	18
II.4.2 Application de la loi de Zipf inverse en analyse d'image	19
II.4.3 Codages appliqués aux images pour leur analyse avec les lois de Zipf et de Zipf inverse..	19
II.4.4 Représentation graphique des lois de Zipf et de Zipf inverse.....	21
II.5 Conclusion.....	24

Chapitre III : Dépistage du cancer du sein

III.1 Introduction.....	25
III.2 Le cancer du sein.....	26
III.3 Les pathologies mammaires.....	27
III.4 Le dépistage du cancer du sein.....	30
III.5 L'imagerie médicale.....	31
III.6 La mammographie.....	31
III.6.1 Les types de mammographies.....	33
III.6.2 Les avantages de la mammographie numérique.....	34
III.7.Conclusion.....	35

Chapitre IV : Aide au diagnostic médical du cancer du sein assisté par ordinateur

IV.1 Introduction.....	36
IV.2 Détection et classification assistées par ordinateur (CAdE/CAdx) des tumeurs dans la mammographie.....	37
IV.2.1 Détection des tumeurs assistée par ordinateur (CAdE).....	38
IV.2.2 Diagnostic des tumeurs assisté par ordinateur (CAdx).....	39
IV.3 Indexation et recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur.....	40
IV.4 L'apport d'un système d'indexation et de recherche de mammographies par le contenu (CBMIIR) par rapport à un système d'aide au diagnostic assisté par ordinateur (CAdx) se basant sur la sortie d'un classifieur.....	42
IV.5 Conclusion.....	44

Chapitre V : Conception et implémentation

V.1. Introduction.....	45
V.2. Problématique.....	46
V.3. Approche proposée basée fusion des lois de puissance : Zipf, Zipf inverse et Haralick pour la classification des zones d'intérêt des mammographies.....	47
V.3.1. Analyse et caractérisation de la texture des zones d'intérêt par les lois de Zipf et de Zipf inverse.....	47
V.3.2. Analyse et caractérisation de la texture des zones d'intérêt par Haralick.....	51
V.3.3. Comparaison entre les lois de Zipf et Zipf inverse et la matrice de co-occurrence pour l'analyse de la texture d'image.....	56
V.4. Classification.....	57
V.4.1. Les méthodes de classification non-supervisées.....	58
V.4.2. Les méthodes de classification supervisées.....	58

V.5 Etapes suivies et accomplies pour la réalisation du système Mammo-Zipf-Har d'indexation et la recherche des mammographies par le contenu.....	59
V.5.1 Caractérisation de la texture par les lois puissance Zipf et Zipf inverse et la matrice de co-occurrence d'Haralick	59
V.5.2 Sélection des descripteurs les plus discriminants par les algorithmes génétiques.....	59
V.5.3 Classification basée Template-matching.....	63
V.6 Architecture du système Mammo-Zipf-Har.....	65
V.7 Implémentation	67
V.7.1 L'environnement de développement.....	67
V.7.2 La base des images mammaires MIAS.....	67
V.7.3 Evaluation des performances.....	67
V.8 Présentation de quelques vues du système «Mammo-Zipf-Har ».....	71
V.9. Conclusion.....	75
Conclusion générale.....	76
Bibliographie.....	78

LISTE DES FIGURES

Chapitre I Analyse d'image et vision par ordinateur

- Figure I.1** Les niveaux d'analyse d'une image..... 5
- Figure I.2** Illustration de l'histogramme typique d'une mammographie.....6

Chapitre II : les lois puissance

- Figure II.1** Représentation d'une loi puissance dans un repère linéaire.....15
- Figure II.2** Exemple de la loi Mandelbrot.....17
- Figure II.3** Motif original (a) et son codage avec la méthode des neuf classes
(b).....20
- Figure II.4** Motif original (a) et son codage avec la méthode des rangs généraux
(b).....21
- Figure II.5** Courbe de Zipf d'une image.....22
- Figure II.6** Courbe de Zipf inverse d'une image.....23

Chapitre III : Dépistage du cancer du sein

- Figure III.1** Les différents types de microcalcification.....29
- Figure III.2** Equipement de la réalisation d'une mammographie.....32
- Figure III.3** Positionnement du sein durant la génération d'une mammographie.....32
- Figure III.4** Processus de réalisation de la mammographie numérique.....34

Chapitre IV : Aide au diagnostic médical du cancer du sein assisté par ordinateur

Figure IV.1 Différence entre les systèmes CADe et les systèmes CADx.....	40
Figure IV.2 Architecture d'un système d'indexation et de recherche des images par le contenu.....	41

Chapitre V : Conception et implémentation

Figure V.1 Courbes de Zipf et de Zipf inverse de trois zones d'intérêts encodées par le codage des rangs généraux : N zone d'intérêt sans tumeur, B zone d'intérêt portant une tumeur bénigne, et M zone d'intérêt portant une tumeur maligne.....	48
Figure V.2 : Plus proches voisins du pixel 'x' selon 4 directions.....	51
Figure V.3 Les quatre directions pour obtenir la matrice cooccurrence.....	52
Figure V.4 Exemple de matrices de cooccurrence construites à partir d'une image 4 × 4 composée de 4 niveaux de gris.....	53
Figure V.5. Fonctionnement général d'un AG de base.....	62
Figure V.6. Sélection de caractéristiques par un algorithme génétique.....	63
Figure V.7 Opérateur de croisement	63
Figure V.8 Exemple d'une opération de mutation.....	63
Figure V.9 Résultat d'algorithme génétique.....	63
Figure V.10 Architecture global du système Mammo-Zipf-Har proposé.....	66
Figure V.11 Fenêtre Login du système «Mammo-Zipf-Har».....	71
Figure V.12 Interface principale du système «Mammo-Zipf-Har».....	72
Figure IV.13 Processus de décision médicale issue de l'utilisation de la sortie de Mammo_Zip-Har par la considération de k=2.....	73
Figure IV.14 Processus de décision médicale issue de l'utilisation de la sortie de Mammo-Zipf-Har par la considération de k=5.....	74

LISTE DES TABLEAUX

Chapitre I Analyse d'image et vision par ordinateur

Tableau II.1 Classification des niveaux de gris en 9 classes.....	20
--	----

Chapitre V : Conception et implémentation

Tableau V.1 Illustration d'une comparaison entre l'approche d'Haralick et celle des Zipf et Zipf inverse pour l'analyse de la texture d'image.....	57
---	----

Tableau V.2 Variation de la précision de classification du système «Mammo-Zipf-Har» sans la réduction des vecteurs descripteurs par les algorithmes génétiques.....	68
--	----

Table V.3 Performance du système «Mammo_Zip-Har» avec et sans Algorithme génétique.....	69
--	----

LISTE DES ACRONYMES

ACP	Analyse en Composantes Principales
CADe	Computer Aided Detection
CADx	Computer Aided Diagnosis
CIRC	Centre International de Recherche sur le Cancer
CBMIIR	Content Based Mammogram Image Indexing and Retrieval
CBIR	Content Based Image Retrieval
INSP	Institut National de Santé Publique
IRM	Imagerie par Résonance Magnétique
K-PPV	K- Plus Proches Voisins
K-NN	K-Nearest Neighbor
LBP	Local Binary Pattern
MIAS	Mammographic Image Analysis Society
MATLAB	MATrix LABoratory
PNN	réseau de neurones probabiliste
ROI	Region Of Interest
TEMP	Tomographie par Emission Monophotonique
SFTA	Segmentation-based Fractal Texture Analysis
SVM	Support Vector Machine
SGLD	(Spatial Gray Level Dependency)
TFD	Transformée de Fourier Discrète
TCD	Transformée de Cosinus Discrète

INTRODUCTION GÉNÉRALE

Le cancer du sein est la forme de cancer la plus répandue chez les femmes du monde entier, seule sa détection précoce peut augmenter la capacité du traitement et le taux de survie des patientes. En effet, ceci est abouti à travers la mammographie de dépistage.

Récemment, des systèmes de détection ou de diagnostic assistés par ordinateur (CAD) ont été développés pour aider les radiologistes à accroître la précision du diagnostic en leur fournissant un deuxième avis, l'aidant à établir le diagnostic final. Effectivement, il est connu que l'interprétation de la mammographie désigne une tâche répétitive qui nécessite une attention maximale afin d'éviter l'interprétation erronée, à cet effet, l'apparition des systèmes de détection ou de diagnostic assisté par ordinateur (CAD) se basant sur des techniques de traitement d'image et de reconnaissance de les formes a eu lieu.

La problématique tirée et discutée pour l'élaboration de ce travail consiste en l'analyse des images mammaires par des approches non linéaires qui répondent mieux à la structure complexe que présente une image au lieu d'appliquer de simples méthodes linéaires qui montrent vite leurs limites.

Nous nous intéressons dans notre cas d'étude aux lois puissance : Zipf et Zipf inverse pour l'analyse de la texture des images mammaires. Plus précisément, nous proposons une fusion des lois de Zipf et de Zipf inverse avec l'approche d'analyse de texture d'Haralick dans le but d'explorer l'apport complémentaire que peut générer ces deux approches de caractérisation de la texture afin d'essayer d'améliorer les résultats obtenus par l'application des lois de Zipf et de Zipf inverse individuellement.

L'objectif de notre travail est la réalisation d'un système fiable et performant d'indexation et de recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein. En effet, l'ultime but est de renforcer la performance du diagnostic assisté par ordinateur (CADx) basé sur un résultat numérique donné par un classifieur dont les radiologues

Introduction générale

jugent ambigus et difficile à interpréter, par un diagnostic à base de cas traitées et stockés préalablement dans une base de données.

Nous avons structuré notre mémoire comme suit :

Chapitre I : Analyse d'image et vision par ordinateur

Dans le chapitre I, nous présenterons les notions et concepts de base de l'analyse d'image et de la vision par ordinateur. Ensuite nous détaillerons l'analyse de la texture dans les images.

Chapitre II : Les lois puissance

Dans le chapitre II, nous présenterons un ensemble de lois puissances en se focalisant sur deux lois : Zipf et Zipf inverse en analyse d'image.

Chapitre III : Dépistage du cancer du sein

Nous introduirons dans le troisième chapitre, le cancer du sein et l'importance accordée à ce type de cancer. Notamment, la mammographie désignant un outil incontournable pour le dépistage du cancer du sein.

Chapitre IV : Aide au diagnostic médical du cancer du sein assisté par ordinateur

Au long du quatrième chapitre, les éléments fondamentaux d'un système de diagnostic assisté par ordinateur sont présentés.

Chapitre V : Conception et Implémentation

Le chapitre V sera consacré à la présentation de notre conception pour le développement du système d'indexation et de recherche des mammographies par le contenu ainsi que les résultats obtenus.

Nous achèverons ce travail par une conclusion générale qui résumera les pivots de notre contribution tout en posant les perspectives tirés.

I.1. Introduction

Dans le traitement et l'analyse des images, la relation spatiale entre les pixels permet de mieux comprendre le contenu de l'image. Ainsi, une approche basée sur le contenu d'image peut être adoptée pour une analyse fructueuse de cette dernière. En effet, nous nous intéressons à l'analyse de la texture d'une image.

La texture est un attribut visuel important dans la vision par ordinateur et a fait l'objet de diverses applications citons la télédétection, l'analyse d'images médicales ainsi que la reconnaissance faciale. L'enjeu majeur consiste en le fait qu'il est facile pour le système visuel humain de distinguer les modèles de texture tandis que ceci est un grand défi pour les méthodes automatiques.

Le but principal des approches d'analyse de texture consiste à extraire un certain nombre de caractéristiques et à les exprimer sous une forme limitée. L'étape d'extraction des paramètres précède souvent l'étape de prise de décision afin que nous puissions répondre à des questions telles que: des tissus biologiques sains ou pathologiques?

Nous évoquons dans le premier chapitre les notions et concepts de base de l'analyse d'image et de la vision par ordinateur. Entre autre, vu que nous nous focalisons sur l'importance de l'analyse numérisée de la texture des images pour la prise de décision, nous donnons un aperçu sur les différentes approches proposées dans la littérature.

I.2. Concept de l'analyse d'image

Les images sont des matrices d'un ensemble de pixels caractérisés par leur intensité lumineuse variable, représentant une scène concrète ou abstraite.

Le terme analyse d'image désigne l'analyse automatique des images numérisées pour en extraire des mesures saillantes. Actuellement, l'analyse d'image comprend de nombreuses classes d'outils, tels que le traitement du signal et l'intelligence artificielle.

Une image numérique est une grille rectangulaire d'échantillonnage dont les constituants sont des pixels portant des informations sur l'intensité lumineuse des différents lieux au sein de l'image. De ce fait, une image numérique est représentée par une matrice bidimensionnelle, dont les éléments sont des nombres naturels correspondant à des niveaux de quantification dans l'échelle de l'intensité lumineuse.

Tout d'abord une image d'entrée est capturée par une caméra de télévision en (2D) et numérisée, étant décrite par une fonction d'image $f(x,y)$ dont la valeur est l'intensité lumineuse en fonction de deux paramètres x et y , représentant les coordonnées de l'emplacement dans l'image. [1][2]

En vision artificielle, nous distinguons le traitement d'images réagissant après les étapes d'acquisition et de numérisation pour assurer les transformations d'images ainsi que la phase de calcul qui permet la réalisation d'une interprétation des images traitées. Effectivement, l'appel à l'intelligence artificielle est fondamental pour la manipulation des connaissances sur les informations dont nous disposons à propos de ce que représentent les images traitées.

Si nous nous interrogeons sur le rôle du traitement d'image, nous pouvons citer l'exemple de la radiographie où ce dernier est essentiellement utilisé pour avoir une meilleure qualité et visibilité des organes sur l'image et permettre ainsi de lever des indéterminations. En effet, de nos jours les médecins sont confrontés à l'utilisation d'ordinateurs pour l'élaboration d'un diagnostic complet lors de la suspicion d'une anomalie. [3][4]

I.2.1 Les différentes phases de l'analyse d'une image

Nous présentons sur la figure I.1 les niveaux d'analyse d'une image où nous distinguons une analyse de bas niveau, une autre de haut niveau et une analyse intermédiaire.

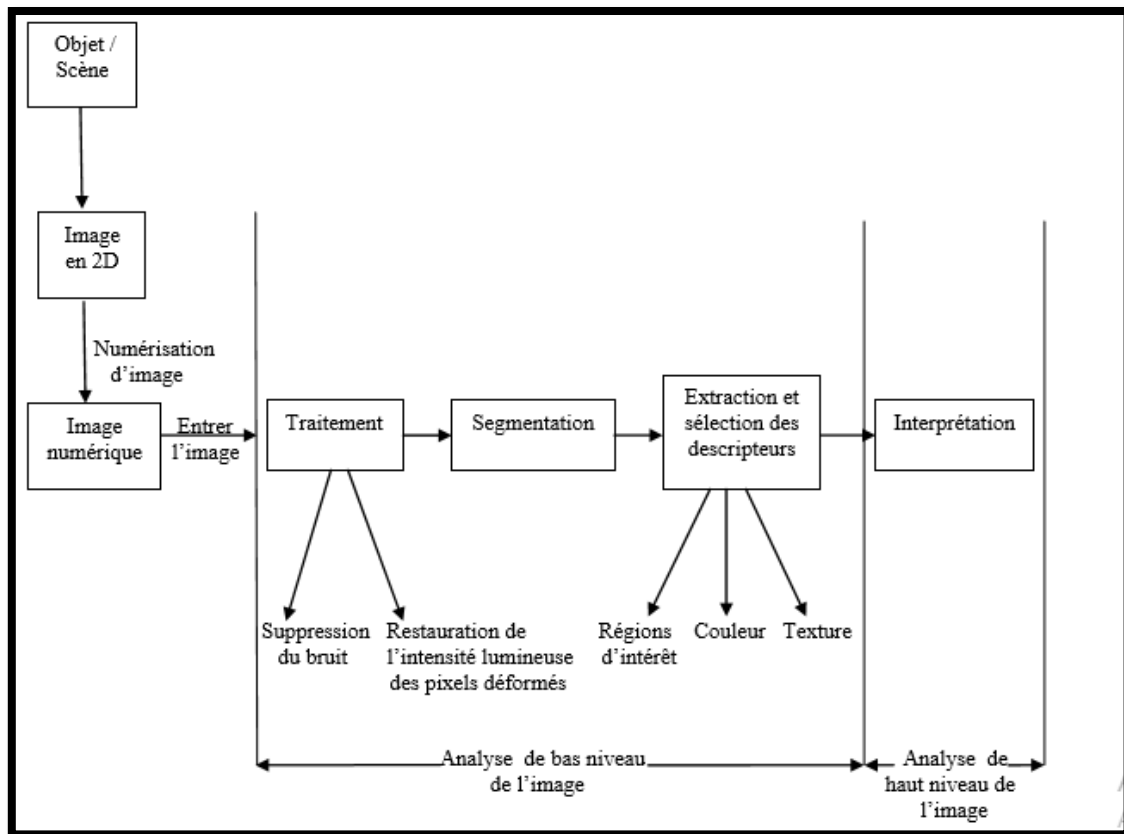


Figure I.1 Les niveaux d'analyse d'une image. Source [5]

- **Analyse de bas niveau d'image**

L'analyse bas niveau consiste en les techniques de numérisation de l'image en nécessitant très peu de connaissances sur le contenu des images. Dans ce sens, l'échantillonnage désigne le procédé de discrétisation spatiale d'une image consistant à associer à chaque zone rectangulaire $R(x,y)$ d'une image continue une unique valeur $I(x,y)$ pour la résolution spatiale. Tandis que la quantification désigne le choix du bon nombre de bits pour coder les images numériques en affectant une valeur numérique à chaque échantillon prélevé sur le signal tout en réduisant le nombre de bits nécessaires au codage.[5][6]

Les buts de la numérisation des images consistent en : visualiser les images sur un moniteur, les imprimer, les traiter sur un ordinateur, les stocker sur des supports informatiques ou les transmettre sur des réseaux informatiques. [5][6].

Nous pouvons citer le traçage des histogrammes d'une image comme une analyse de bas niveau, en effet, un histogramme est un graphique statistique permettant de représenter la distribution des intensités des pixels d'une image en extrayant le nombre de pixels pour chaque intensité lumineuse. Entre autre, le seuillage désigne à son tour une analyse de bas

niveau d'une image en considérant uniquement les intensités des pixels et non pas les relations de voisinage entre eux. Par exemple pour un seuil donné S , le compromis régions/pixels doit toujours être respecté. En fait l'ultime but est de trouver un seuil optimal souvent choisi par tâtonnement devant minimiser les faux contours (donc il ne doit pas être trop bas) tout en conservant le maximum de régions complètes (donc il ne doit pas être trop haut). [7][5]

La figure I.2 illustre l'historgramme typique d'une mammographie où nous distinguons trois classes différentes, à savoir, classe de fond, classe du sein et classe regroupant les annotations, bande d'intensité et les calcifications.

Classe1: située sur la gauche, elle regroupe tous les pixels ayant des valeurs de faible intensité, le grand pic correspond aux pixels de fond.

Classe2: située au milieu, elle regroupe tous les pixels ayant des valeurs grises correspondant au sein lui-même.

Classe3: située sur la droite, elle regroupe tous les pixels de haute intensité, le grand pic correspondant au muscle pectoral, aux annotations ainsi qu'à des lésions mammaires [7]

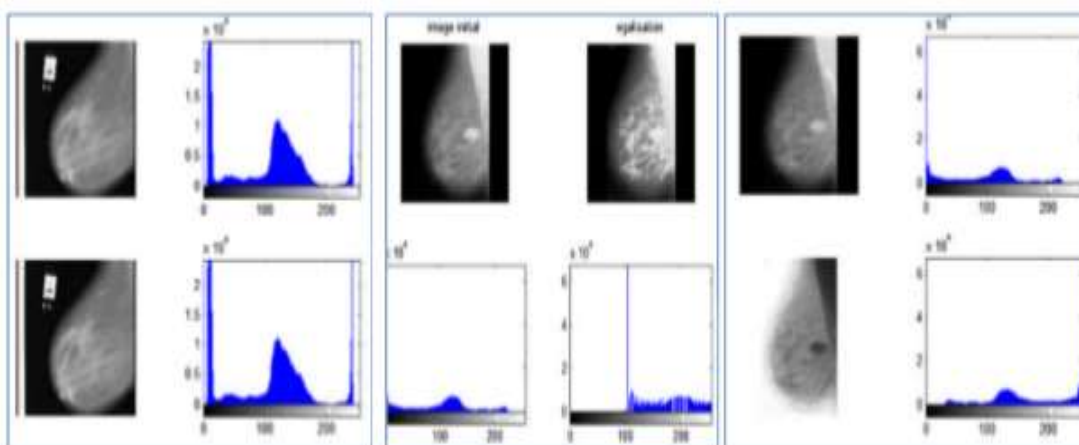
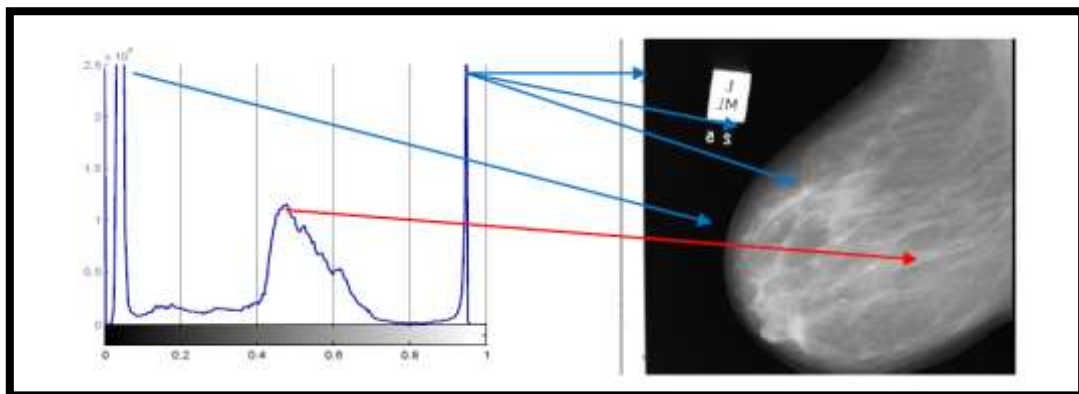


Figure I.2 Illustration de l'historgramme typique d'une mammographie. Source [8]

- **Analyse d'intermédiaire niveau d'image**

Le niveau intermédiaire désigne le passage d'une représentation purement basée sur l'image à une représentation d'objet. En effet, l'enjeu est l'identification des régions saillantes à l'analyse de haut niveau. Dans ce sens, différents algorithmes de segmentation en régions et des contours, ainsi que des algorithmes d'extraction de caractéristiques ou de descripteurs sont utilisés [6].

- **Analyse de haut niveau d'image**

La vision de haut niveau s'appuie sur une certaine forme du modèle formel du monde, par la suite, la réalité perçue sous la forme d'images numériques est comparée au modèle; l'ordinateur opte pour les résultats atteints de l'analyse d'image à bas niveau pour mettre à jour le modèle [10]. Notons qu'à ce niveau, les méthodes d'intelligence artificielle sont largement applicables. [5]

I.3 Concept de la vision par ordinateur

La vision par ordinateur se perçoit comme une initiative dans l'automatisation et l'intégration d'une large gamme de processus sur la perception visuelle.

Dans le domaine du traitement de l'image, la vision par ordinateur a pris une part de plus en plus importante. En effet, l'analyse de l'imagerie médicale constitue l'outil principal de la vision par ordinateur tout en distinguant d'autres champs applicatifs comme la télésurveillance, la télédétection ou la biométrie.

Suite à l'apparition de la vision par ordinateur, l'analyse d'image ne se limite plus à segmenter une image en régions uniquement ou à en extraire les contours, l'ordinateur se charge d'exploiter ces informations pour donner un sens aux images. Ceci en occurrence avec le fait d'agir de la même façon que l'homme lorsqu'il interprète son environnement grâce à sa vue. [11]

La vision par ordinateur a pour but de trouver une relation entre une image d'entrée et les modèles du monde réel. À cet effet, le passage d'une image à un modèle rend oblige l'utilisation des connaissances sémantiques sur l'interprétation des données d'images.

Notons qu'en vision par ordinateur, les images sont acquises par une caméra puis subissent la numérisation à travers un ordinateur via une carte d'acquisition. [12]

Dans la conception moderne de l'Intelligence Artificielle, la machine perçoit son

environnement et s'y adapte. Le traitement d'image ainsi que la vision s'inspirent des connaissances et des techniques d'intelligence artificielle pour gérer la prise de décision. [13]

I.4 Analyse de la texture d'image

L'analyse de la texture demeure une problématique ouverte et une tâche majeure dans de nombreux domaines de la vision par ordinateur et de la reconnaissance de formes, y compris l'imagerie médicale.

I.4.1 Définition de la texture

Donner une définition précise et universelle du concept de texture demeure un problème ardu sans résolution car il n'existe pas un véritable consensus pour la définition d'une texture. Cependant, une judicieuse compréhension de la notion de texture permettrait d'aborder une analyse plus méthodique. [14]

Dans le traitement d'images, l'analyse de texture joue un rôle très important consistant en l'identification des objets et la texture est considérée comme une région d'une image présentant une organisation spatiale homogène des niveaux de luminance [15]. En outre, une texture peut être vue comme un champ de l'image apparaissant sous forme d'un domaine cohérent et homogène, donc, formant un tout pour un observateur [16]. Ou bien encore, cette définition [17] qui considère une texture comme un phénomène à 2 dimensions:

- Une première concernant la description d'éléments de base ou primitives, à partir desquels la texture est formée.
- La deuxième est relative à la description des relations spatiales entre ces primitives

Dans [18], la texture est définie comme suit : « Une texture est un champ de l'image qui apparaît comme un domaine cohérent et homogène, c'est-à-dire formant un tout pour un observateur » [19][27].

Haralick [14] a décrit la notion de texture comme suit : une image texturée est distinguée par le nombre et les types de ses primitives tonales ainsi que leurs orientations spatiales. Une texture est une région d'une image pour laquelle nous définissons une fenêtre de dimensions minimales, telle qu'une observation à travers de celle-ci se traduit par une perception visuelle identique pour toutes les translations possibles de cette fenêtre à l'intérieur de la région considérée [14].

I.4.2 Caractérisation de la texture d'image

Il existe de nombreuses méthodes pour analyser les textures. Nous regroupons généralement ces approches en quatre familles : [19]

- **Les approches statistiques:** citons la fonction d'autocorrélation, les matrices de co-occurrences.
- **Les approches Spatio-fréquentielles:** citons la transformée de Fourier, les bancs de filtres, la transformée en ondelettes
- **Les approches basées sur des modèles:** citons les Champs de Markov.
- **Les approches structurelles :** citons les composantes connexes, local binary pattern. [9].

Nous allons détailler ces approches comme suit [9] [18]:

❖ Les méthodes statistiques

- **Les statistiques de premier ordre (basées sur l'histogramme)**

Elles sont basées sur la distribution de niveau de gris sans la considération de leur distribution relative. Les statistiques du premier ordre sont calculées par rapport à l'histogramme des intensités. Elles présentent l'inconvénient de n'exploiter que l'information de niveau de gris.

- **La matrice de co-occurrence (basées sur les couples de pixels)**

C'est une méthode statistique du second ordre qui est largement utilisée pour le traitement des images médicales [9]. Elle est appelée aussi méthode de dépendance spatiale des niveaux de gris : SGLD. Une matrice de co-occurrence mesure la probabilité d'apparition $P(i, j, \delta, \theta)$ dénombrant le nombre de fois qu'un pixel de niveau d'intensité i apparaît à une distance relative δ d'un pixel de niveau d'intensité j en considérant une orientation θ donnée. [25][26]

Généralement, nous utilisons 0, 45, 90 et 135 degrés pour les directions angulaires θ .

Si nous analysons les caractéristiques extraites à partir de ces matrices, nous affirmons l'obtention des informations sur l'homogénéité, les dépendances linéaires entre les niveaux de gris, le contraste ainsi que la complexité de l'image. [19] [29]

❖ Les méthodes fréquentielle (Spatio fréquentielle)

Ces approches convertissent l'image originale à une nouvelle image en prenant en considération les propriétés des fréquences spatiales de l'intensité des pixels. Ceci peut être appliqué au moyen de Transformée de Fourier Discrète (TFD) bidimensionnelle, ou bien par

la transformée de Cosinus Discrète (TCD), pour un ultime but consistant en l'extraction des composantes de fréquence spatiale des images. [24]

❖ Les méthodes basées sur un modèle

Ces méthodes affirment que les paramètres estimés des modèles a priori prennent le rôle des descripteurs de texture durant la caractérisation de la texture.

• Les modèles autorégressifs

Ils considèrent une interaction locale directe entre les pixels de l'image, de façon que l'intensité des pixels soit une somme pondérée d'intensités de pixels dans un voisinage du pixel et un bruit distribué de manière identique. En effet, les paramètres du modèle consistent en le vecteur des pondérations. [23][19]

• Les champs aléatoires

Ils supposent que la texture est considérée comme un échantillon fini d'un processus aléatoire bidimensionnel pouvant être décrit par ses paramètres statistiques. [9]

• Fractales

Une fractale désigne un concept mathématique dans lequel un ensemble présentant plusieurs échelles est le même motif répétitif à chaque échelle, ce qui est un paradigme qui a été transféré à l'analyse de texture. Notons que les paramètres fractals désignent une mesure d'irrégularité ou d'hétérogénéité des arrangements spatiaux. [19]

❖ Les méthodes structurelles

Elles sont basées sur les textures composées des motifs (ou bien primitives) qui se répètent fréquemment et régulièrement dans l'espace, tout en tentant de décrire les règles qui régissent l'organisation spatiale de ces motifs. [5]. [21]

Notons que les méthodes structurelles sont les meilleurs pour la description de la texture de grande résolution

I.4.3 Domaines d'application de l'analyse de la texture

Si nous ne considérons que l'étude de la distribution des niveaux de gris pour la caractérisation des zones homogènes au sein d'une image, ceci ne sera pas suffisant. En effet, le recours à l'utilisation de la texture a eu lieu pour une judicieuse caractérisation, qui sera suivie de reconnaissance et d'interprétation ultérieure de l'image. En effet, l'analyse de la texture est devenue de plus en plus fréquente dans divers domaines [26] :

- **L'imagerie médicale**

Des descripteurs discriminant de texture permettent de faire la distinction entre tissus sains et pathologiques. Citons la détection de lésions, le dépistage de pathologie, la segmentation en IRM (Image Résonance Magnétique) ou bien en mammographie.

- **La Géologie**

En analysant l'état du sol ainsi que les données sismiques et volcaniques.

- **L'imagerie satellitaire**

En analysant et interprétant les images satellitaires, en effet, une estimation discriminante de la texture permettra de différencier un champ d'une forêt par exemple. Nous citons un exemple récent qui a permis aux archéologues de reproduire les chemins que parcoururent les caravanes de commerce dans le moyen orient [22].

- **La Robotique**

Nous illustrons l'exemple de la navigation d'un robot mobile. En effet, en explorant la texture de la surface nous pourrions avoir des informations sur l'environnement de la navigation tout en récoltant des données sur la forme, la profondeur, l'orientation ainsi que la distance des surfaces visibles.

- **La compression d'images**

Les descripteurs texturaux permettent de représenter l'image sous forme d'un nombre optimal de paramètres évidemment, la reconstruction de l'information s'effectuera selon un minimum d'erreurs.

- **Contrôle non destructif**

L'analyse de texture a fait l'objet d'une large utilisation dans l'inspection des surfaces en contrôle de qualité (produits industriels, matériaux, produits alimentaires, détection de défauts).

I.5 Conclusion

Au cours de ce chapitre nous avons évoqué que la principale difficulté de l'analyse tissulaire désigne l'absence d'une définition universelle de la texture. En effet, le terme texture obéit à des acceptions différentes selon le domaine dans lequel il est utilisé (géologie, pétrologie, textile, agro-alimentaire).

Dans l'analyse d'image, la texture demeure l'élément le plus important des informations fournies sur l'image. Dans ce sens, diverses méthodes d'analyse de la texture ont été proposées dans la littérature ce qui engendre l'apparition de plusieurs descripteurs de texture. En effet, par exemple le spectre de puissance de la transformé de Fourier est très informatif sur la périodicité d'une texture. Un autre moyen d'acquérir des descripteurs de texture d'image consiste à extraire l'information à partir de la matrice de co-occurrence de niveau de gris, également appelée matrice de dépendance spatiale de niveau de gris [24].

Pour faire face aux difficultés d'analyse des textures, notre problématique consiste en la réalisation d'une caractérisation judicieuse de la texture des images mammaires. En effet, nous allons proposer une fusion de l'approche d'Haralick qui semble plus appropriée, avec celle des lois puissance Zipf et Zipf inverse pour la caractérisation de la texture des zones d'intérêt extraites à partir d'images mammaires dans un processus d'aide au diagnostic médical du cancer du sein assisté par ordinateur.

Le chapitre suivant portera sur les modèles de loi puissance de type: Zipf et Zipf inverse pour donner un aperçu sur leur application dans le domaine de l'imagerie.

II.1. Introduction

A travers ce chapitre nous allons montrer la contribution de différentes lois puissance dans la modélisation des phénomènes extrêmement divers tels que les lois de Zipf et de Zipf inverse qui établissent respectivement, une relation entre les fréquences d'apparition des différentes occurrences d'un phénomène et le rang de ces occurrences dans une suite ordonnée, ainsi qu'une relation entre les fréquences d'apparitions des attributs et le nombre d'attributs ayant la même fréquence d'apparition.

Nous exposons par la suite les principales applications de diverses lois puissance existantes, tout en détaillant l'application des lois puissance dont nous nous intéressons tout au long de notre projet, à savoir, la loi de Zipf et de Zipf inverse dans le domaine de l'imagerie.

II.2. Qu'est-ce qu'une loi puissance ?

On peut définir une loi puissance comme étant une fonction associant deux quantités. Elle peut être de la forme suivante $y = ax^{-b}$ avec a et b des constantes. Une représentation graphique dans un repère $(x ; y)$ est la suivante :

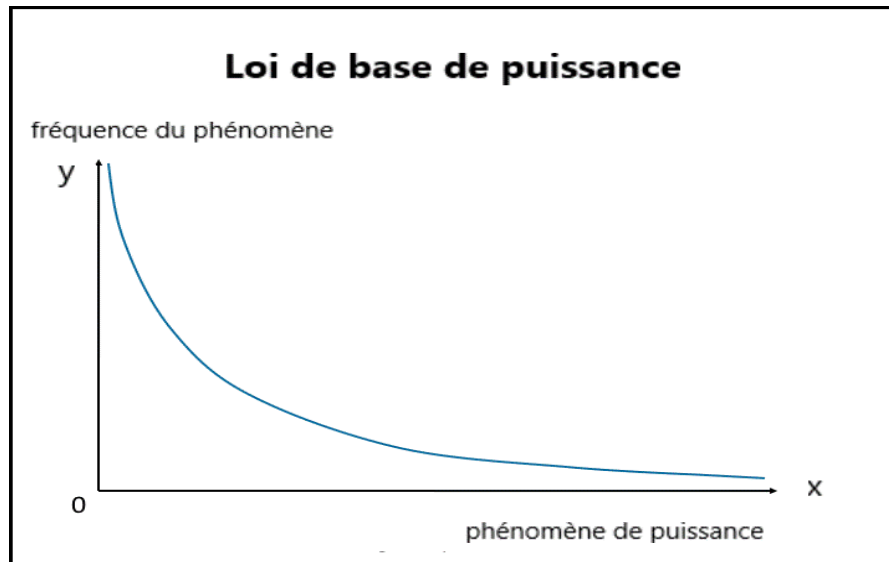


Figure II.1 Représentation d'une loi puissance dans un repère linéaire. Source [6]

Généralement, les lois puissance sont représentées dans un repère bi-logarithme. En choisissant un tel repère, nous obtenons la représentation graphique d'une distribution en loi puissance sous la forme d'une droite. Nous aurons donc $\ln y = \ln a - b \ln x$

Les modèles de loi puissance les plus connus sont la loi de Zipf et la loi de Pareto. Ils sont omniprésentes dans différents domaines scientifiques (physique, biologie, psychologie, sociologie, économie, linguistique). [30]

II.3. Les lois puissance existantes

II.3.1 La loi de Pareto

Selon [36], la loi de Pareto découle d'une démarche inductive cherchant à utiliser des outils statistiques existants pour explorer la régularité de certains phénomènes économiques et notamment la distribution des revenus.

Chapitre II : Les lois puissance

Le principe de Pareto est très utile pour prendre certaines décisions, en effet, dans le domaine de la conception d'interface, il suppose une analyse approfondie de l'expérience utilisateur. Nous pouvons constater que ce principe est applicable toujours où il y a des données en nombre assez suffisant pour être quantifié.

La loi de Pareto, aussi appelée loi des 80/20, est une loi empirique inspirée par les observations suivantes de Pareto: environ 80 % des effets est le produit de 20 % des causes, si nous étudions la répartition des revenus personnels des individus dans les principaux pays industrialisés nous affirmerons que 80% des richesses appartiennent à 20% de la population. Dans ce sens, Pareto a constaté que cette répartition suivait une loi puissance. [5]

II.3.2. Loi de Benoit Mandelbrot

Les gens ont toujours pris plaisir à concevoir des «problèmes» dans le but de poser un défi ou d'offrir un plaisir intellectuel. Ainsi, de nombreuses récréations mathématiques d'origine qui réapparaissent de temps en temps sous de nouveaux habits semblent avoir survécu principalement parce qu'elles font appel au sens de la curiosité ou au mystère de l'homme.

Les problèmes en général étaient de deux types: ceux impliquant la manipulation d'objets et ceux nécessitant un calcul. La première nécessitait peu ou pas de compétences en mathématiques, mais simplement d'intelligence générale et d'ingéniosité, comme par exemple les problèmes de traversée et de traversée difficiles. Un exemple typique du premier est comment mesurer un quart de liquide si on ne dispose que de mesures de huit, cinq et trois pintes. Comme exemple :

Pensez à un nombre, triplez-le et prenez la moitié du produit; triple ceci et prends la moitié du résultat; divisez ensuite par 9. Le quotient sera le quart du nombre initial.

Certains groupements de nombres naturels, opérés par les processus ordinaires de l'arithmétique, révèlent des modèles assez remarquables, offrant des passe-temps agréables. Comme mentionné sur la figure II.2 [31] :

$1 \times 8 + 1 = 9$	$3 \times 37 = 111$	$(1)^2 = 1$
$12 \times 8 + 2 = 98$	$6 \times 37 = 222$	$(11)^2 = 121$
$123 \times 8 + 3 = 987$	$9 \times 37 = 333$	$(111)^2 = 12321$
$1234 \times 8 + 4 = 9876$	$12 \times 37 = 444$	$(1111)^2 = 1234321$
etc.	etc.	etc.

Figure II.2 Exemple de la loi Mandelbrot. Source [31]

II.3.3. La loi de Robert GIBRAT

Cette loi trouve son origine au début du 20^{ème} siècle. En effet, en explorant un ensemble de données portant sur un large panel issu du panel de la Fondation ZEW (Ouest), la loi de Gibrat affirme que la croissance en pourcentage des entreprises est indépendante de la taille initiale de l'entreprise[31], en affirmant que les plus petites entreprises croissent plus vite en comparaison avec leurs homologues possédant une plus grande taille.[31]

II.3.4. La loi de Zipf

La loi de Zipf est une loi empirique énoncée en 1949 par G.K Zipf [32]. La loi de Zipf se base sur la répartition statistique des fréquences d'apparition des mots appartenant à un langage. Selon la loi de Zipf, les fréquences d'apparition des mots classés par ordre décroissant de leur fréquence d'apparition s'organisent suivant une loi puissance. [30]

En effet, si nous désignons par F la fréquence d'apparition du mot de rang r dans la suite, la relation est donnée par la formule suivante [32] :

$$F = a r^{-\alpha} \quad (\text{II.1})$$

Dans cette formule, a et α sont des constantes et la valeur de l'exposant α est proche de 2 pour les textes écrits en langage naturel et caractérise la loi puissance.

Cette distribution en loi puissance se représente graphiquement en échelle bi-logarithmique, avec en abscisse le rang des motifs et en ordonnée leurs fréquences d'apparition [32].

II.3.5 La loi de Zipf inverse

La loi de Zipf inverse décrit la répartition statistique des fréquences des mots. Selon la loi de Zipf inverse, le nombre de mots distincts ayant une fréquence d'apparition donnée dans un texte suit une loi puissance.

Le nombre $I(f)$ de mots apparaissant f fois dans le texte est donné par la formule suivante [30] [5] :

$$I(f) = a f^{-\gamma} \quad (\text{II.2})$$

Dans cette formule, a et γ sont des constantes.

II.4 Application des lois de Zipf et de Zipf inverse en analyse d'image

Dans le domaine de l'analyse d'images, la loi de Zipf a déjà été utilisée avec succès par [30] pour la mesure de la qualité des images compressées et par [30] pour la détection des zones d'intérêt.

Une loi puissance permet de modéliser certains phénomènes avec une approximation précise et dont il est difficile de concrétiser au moyen de modèles linéaires. Ceci justifie l'application de ces modèles dans le cas des images, en effet, il est connu que l'image possède une structure complexe pour être approximée par de simples lois linéaires [33].

Evidemment, pour l'analyse d'image par les lois de Zipf et de Zipf inverse nous travaillerons avec des matrices d'un ensemble de pixels caractérisés par leur intensité lumineuse variable. [35]

II.4.1. Application de la loi de Zipf en analyse d'image

Pour analyser une image avec la loi de Zipf, nous parcourons l'image avec un masque de dimension $m * m$, par la suite, les motifs rencontrés sont codés en utilisant un codage appliqué aux pixels de l'image.

Nous classons ces motifs suivant l'ordre des fréquences décroissantes et nous attribuons le rang de chaque fréquence. La courbe rang fréquence est ensuite tracée dans un repère bi-logarithmique.

Nous pouvons modéliser cette relation par la formule suivante [30] :

$$N_{\sigma(i)} = k i^{-a} \quad (\text{II.3})$$

Où $N_{\sigma(i)}$ représente la fréquence d'apparitions d'un motif de rang i , k et a sont des constantes.

Si nous analysons l'adaptation faite pour le passage du texte à l'image, nous affirmerons que les symboles considérés pour l'application des lois de Zipf sont les niveaux de gris utilisés pour coder les pixels. En effet, les n -uplets sont considérés comme la suite des niveaux de gris des pixels constituant des masques susceptibles de prendre des formes variées : des masques carrés 3×3 ou des masques linéaires verticaux 3×1 , 8×1 ou encore horizontaux, 1×3 , 1×8 . Notons que ce choix dépend du type de motif que nous souhaitons rechercher, à savoir : motif linéaire ou motif surfacique. Notons que d'autres tailles de motifs seraient possibles, tandis qu'une taille plus grande des motifs aurait pour conséquence le fait que chaque motif n'aurait qu'une minime probabilité de se retrouver plusieurs fois dans l'image, et la distribution des fréquences des motifs n'aurait pas une véritable signification. Il est utile de mentionner aussi qu'une taille de motifs relativement importante augmentera considérablement le temps de calcul [33].

II.4.2 Application de la loi de Zipf inverse en analyse d'image

Le principe de cette loi consiste à s'intéresser aux motifs distincts qui ont la même fréquence d'apparition. D'après la loi de Zipf inverse, le nombre I des motifs différents ayant une fréquence d'apparition f est donné par la formule suivante [30] :

$$I(f) = a f^{-\gamma} \quad (\text{II.4})$$

Où a et γ sont des constantes.

II.4.3 Codages appliqués aux images pour leur analyse avec les lois de Zipf et de Zipf inverse

Le codage de l'image a pour objectif de réduire le nombre de motifs distincts à une valeur qui nous permettra d'en dégager une information pertinente [30] [35]. Donc, nous nous trouvons face à la nécessité de définir un codage aux images.

Chapitre II : Les lois puissance

Nous distinguons deux codages possibles :

- **Codage des 9 Classes**

Ce codage consiste à réduire le nombre des motifs distincts en divisant l'échelle de niveau de gris [0,255] en 9 intervalles, ou bien de classes de largeur égale, évidemment, numérotées en ordre croissant de 0 à 8 comme suit :

Tableau II.1 Classification des niveaux de gris en 9 classes [5].

0-27	28 -55	56 - 83	84 - 111	112 - 139	140 -167	168-195	196-223	224-255
0	1	2	3	4	5	6	7	8

Chaque pixel du motif a une valeur de classe $c(x, y)$ en fonction de la valeur $g(x, y)$ de son niveau de gris, en utilisant la formule suivante [30] [5] :

$$c(x, y) = \text{int} \left[\frac{N * g(x, y)}{255} \right] \quad (\text{II. 5})$$

Nous mentionnons sur la figure II.3 un exemple d'un motif avant et après l'avoir codé par la méthode des 9 classes [5]

(a)

0	0	8
2	0	7
1	4	4

(b)

0	2	250
70	10	213
52	125	120

Figure II.3 Motif original (a) et son codage avec la méthode des neuf classes (b).

- **Codage des rangs généraux**

Le codage par rang généraux est un codage permettant d'étudier la texturation fine de l'image [34] en utilisant un alphabet de taille réduite dans le but de limiter le nombre de motifs possibles.

Par contre le codage des 9 classes permet uniquement de mettre en évidence les principales structures de l'image [30].

Le principe du codage des rangs généraux consiste à remplacer les niveaux de gris des pixels par leur rang dans un voisinage donné. En effet, nous ordonnons les niveaux de gris des pixels du motif selon un ordre croissant et nous affectons la valeur 0 au niveau de gris le plus bas, tout en incrémentant la valeur d'une unité jusqu'à atteindre le niveau de gris le plus élevé.

Il est nécessaire de mentionner que les pixels ayant la même valeur de niveau de gris recevront le même rang.

La figure II.4 présente un exemple d'un motif avant et après son codage par la méthode des rangs généraux.

(a)	(b)																		
<table border="1" style="width: 100%; border-collapse: collapse;"><tr><td style="padding: 5px;">250</td><td style="padding: 5px;">200</td><td style="padding: 5px;">200</td></tr><tr><td style="padding: 5px;">25</td><td style="padding: 5px;">4</td><td style="padding: 5px;">29</td></tr><tr><td style="padding: 5px;">35</td><td style="padding: 5px;">4</td><td style="padding: 5px;">35</td></tr></table>	250	200	200	25	4	29	35	4	35	<table border="1" style="width: 100%; border-collapse: collapse;"><tr><td style="padding: 5px;">5</td><td style="padding: 5px;">4</td><td style="padding: 5px;">4</td></tr><tr><td style="padding: 5px;">1</td><td style="padding: 5px;">0</td><td style="padding: 5px;">2</td></tr><tr><td style="padding: 5px;">3</td><td style="padding: 5px;">0</td><td style="padding: 5px;">3</td></tr></table>	5	4	4	1	0	2	3	0	3
250	200	200																	
25	4	29																	
35	4	35																	
5	4	4																	
1	0	2																	
3	0	3																	

Figure II.4 Motif original (a) et son codage avec la méthode des rangs généraux (b).

II.4.4 Représentation graphique des lois de Zipf et de Zipf inverse

- **Construction de la courbe de Zipf**

La construction de la courbe de Zipf d'une image nécessite l'application d'un algorithme qui consiste à balayer l'image séquentiellement avec un masque de capture de taille 3x3. Puis à réaliser une codification des motifs par un codage adéquats aux propriétés de l'image analysée. [30]

Chapitre II : Les lois puissance

Nous comptons le nombre d'occurrences de chacun des motifs distincts dans l'image. Pour ceci, nous allons ranger les motifs rencontrés dans un tableau. En effet, chaque motif sera comparé par les motifs existant dans le tableau, s'il est présent, nous incrémentons sa fréquence d'apparition d'une unité, autrement, nous l'indexerons dans le tableau en mettant sa fréquence d'apparition à 1.

Une fois l'image est entièrement balayée, les fréquences d'apparition des motifs seront triées dans un ordre décroissant.

Finalement, la fréquence de chaque motif est tracée en fonction de son rang dans un repère bi-logarithmique où l'abscisse représente le rang R des motifs et l'ordonnée représente leur fréquence d'apparition.

Nous présentons sur la figure II.5 la courbe de Zipf relative à une image.

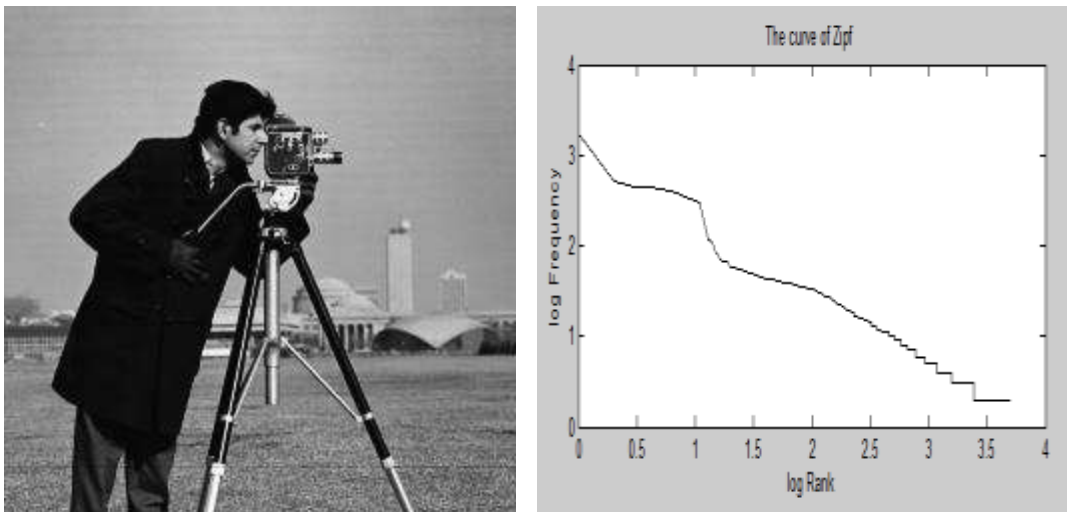


Figure II.5 Courbe de Zipf d'une image. Source [39]

- **Construction de la courbe de Zipf inverse**

L'algorithme de construction de la courbe de Zipf inverse se compose de deux étapes :

La première étape [30] consiste à balayer séquentiellement l'image avec un masque de capture de taille 3x3, ensuite, les motifs seront codés par un codage qui correspond aux propriétés de l'image pour dénombrer les motifs trouvés.

La seconde étape diffère de celle de Zipf, effectivement, nous comptons les motifs ayant la même fréquence d'apparition que la fréquence courante. A cet effet, nous initialisons la

Chapitre II : Les lois puissance

fréquence cherchée à 1 et nous parcourons séquentiellement le tableau des motifs pour compter les motifs de la même fréquence que la fréquence courante.

L'algorithme est réitéré en incrémentant à chaque fois la fréquence d'une unité jusqu'à l'atteinte de la fréquence maximale. Finalement, le nombre de motif est tracé en fonction de leur fréquence d'apparition dans un repère bi-logarithmique constituant la courbe de Zipf inverse.

Nous présentons sur la figure II.6, la courbe de Zipf inverse relative à une image.

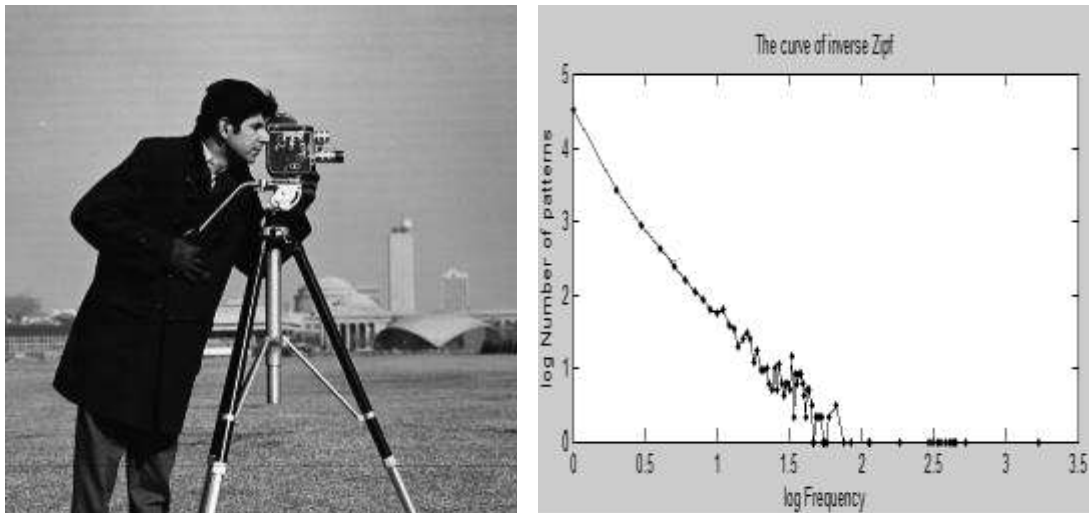


Figure II.6 Courbe de Zipf inverse d'une image. Source [39]

II.5 Conclusion

Les distributions statistiques en loi puissance telle que la loi de Zipf et Zipf inverse sont très communes. Elles ont été mises en évidence, à l'origine, de manière empirique, dans les domaines les plus divers, que ce soit dans des phénomènes naturels ou bien dans des domaines des sciences humaines.

En analyse d'image, les lois de Zipf et de Zipf inverse permettent de caractériser la complexité structurelle d'une image par la répartition statistique des fréquences d'apparition des motifs, en effet, ceci permettra d'analyser la texture de l'image.

Nous allons démontrer dans le prochain chapitre des notions de bases en terme de dépistage du cancer du sein, en effet, notre problématique consiste en l'aide au diagnostic médical du cancer du sein assisté par ordinateur et une étude approfondie concernant les pathologies mammaires doit avoir lieu.

III.1 Introduction

Le cancer du sein est le cancer le plus fréquent touchant la femme. Il représente plus du tiers de l'ensemble des nouveaux cas de cancer chez la femme. Environ une femme sur dix est touchée par cette maladie durant son existence. Particulièrement, le cancer du sein constitue, dans le monde entier, la cause de décès la plus fréquente chez la femme.

Le dépistage précoce favorise les chances de guérison et réduit le taux de mortalité [40]. En effet, nous introduirons dans ce chapitre la notion du cancer du sein, à cet effet, nous devons évoquer brièvement les différentes pathologies mammaires.

Par la suite, nous aborderons les outils du diagnostic du cancer du sein, notamment la mammographie qui est incontournable pour le dépistage du cancer du sein.

III.2 Le cancer du sein

Le cancer du sein est généré par une tumeur associée à la prolifération chaotique et non spécifique de cellules génétiquement modifiées. Cette prolifération provoque la destruction du tissu basal ainsi que l'extension de la tumeur. Malheureusement, le corps ne pourra pas les contrôler.

Selon une analyse de la Société américaine du cancer (ACS) cette maladie pourra causer le décès de 5,5 millions de femmes chaque année dans le monde jusqu'à 2030. La prolifération du cancer en Algérie de la période : 2015 jusqu'à 2018 fait état de 42.700 nouveaux cas par an.

Le cancer du sein commence en premier lieu dans les cellules du sein. En effet, une tumeur cancéreuse ou bien maligne désigne un groupe de cellules cancéreuses pouvant envahir et détruire les tissus voisins par la suite. Voir plus, il peut se propager vers d'autres parties du corps.

Une grande attention a été attribuée pour ce fameux et dangereux cancer. Dans ce contexte, l'institut de surveillance de la santé publique, qui est chargé de surveiller en permanence l'état de santé de la population, a effectué l'étude de l'évolution des décès dus au cancer du sein relative à la population mondiale [40]. Dans cette étude, l'incidence se réfère à la mesure de la fréquence des cas touchés survenant chaque année tout en indiquant le nombre de décès [40].

Différents facteurs peuvent causer le développement du cancer du sein :

- **Antécédents familiaux de cancer du sein et d'autres cancers**
 - Une ou plus d'une parente au premier degré (mère, sœur ou fille) a eu un cancer du sein, spécialement, si le diagnostic a été posé avant la ménopause.
 - Des parentes au deuxième degré (grand-mère, tante ou nièce), paternelles ou maternelles, ont subi un cancer du sein.
 - Une parente s'est fait atteindre du cancer dans les deux seins (bilatéral) avant la ménopause.
 - Deux parentes ou bien plus ont eu un cancer du sein ou celui de l'ovaire.
 - Une parente a eu le cancer du sein et le cancer de l'ovaire.
 - Un parent a eu un cancer du sein.

Il est utile de mentionner que la femme possédant une parente du premier degré subissant un cancer du sein, risque environ deux fois plus d'en être atteinte à son tour. Effectivement, plus

il y a de parentes au premier degré atteintes de ce cancer, plus le risque est élevé. Tandis que le risque n'est pas aussi élevé dans le cas de parentes au second degré.

- **L'âge**

Le facteur de l'âge est le risque le plus important car nous constatons qu'après 40 ans le risque de développer un cancer du sein se multiplie par une fois et demie tous les dix ans [41]

- **L'obésité, notamment après la ménopause**

En effet, une large partie des cancers du sein fut sensibles aux hormones oestrogènes. Et du fait que les tissus graisseux génèrent de petites quantités d'oestrogènes alors si nous avons beaucoup de tissus graisseux, le risque de développer un cancer du sein s'augmentera. [41]

III.3 Les pathologies mammaires

Le corps humain contient trente milliards de cellules portant des informations génétiques nécessaires à la réplication dans le but de remplacer les cellules mortes. Une mutation génétique provoque la prolifération cellulaire sans contrôle ainsi qu'elle favorise l'apparition de nouvelles mutations tout en élevant le rythme des divisions chaotiques. En effet, ces premières cellules cancéreuses forment une tumeur.

Nous distinguons divers types de microcalcifications, à savoir [40] :

- a) **Microcalcifications cutanées ou dermiques**

Elles présentent typiquement un centre clair. Des clichés en incidence tangentielle sont souvent utilisés pour s'assurer de la localisation cutanée de ces microcalcifications.

- b) **Microcalcifications vasculaires**

Ces microcalcifications en rails ou linéaires sont associées à des structures tubulaires.

- c) **Microcalcifications grossières ou coralliformes**

Elles possèdent une grande taille, en effet, supérieures à 2-3 mm de diamètre.

d) Microcalcifications en bâtonnets

Elles sont souvent associées à une ectasie canalaire (dilatation du canal galactophore) et ce fait dirigés vers le mamelon. Leurs mesure habituelle est plus de 1 mm de large et présentent un centre clair si le dépôt calcique se fait dans la paroi du canal.

e) Microcalcifications rondes

Elles présentent une forme ronde avec différentes tailles. En effet, dans le cas où elles mesurent moins de 0.5 mm, elles sont appelées punctiformes ou pulvérulentes.

f) Microcalcifications à centres clairs

Leur taille peut s'étendre de 1 mm à plus de 1 cm. Elles sont rondes ou ovales, à surface lisse et à centre clair. La paroi calcifiée qui les entoure est plus épaisse que celle des microcalcifications en coquille d'œuf.

g) Microcalcifications en coquille d'œuf ou pariétales

Ces microcalcifications très finesse apparaissent sous forme de dépôts calciques sur la surface d'une sphère.

h) Microcalcifications à type de lait calcique

Elles sont sédimentées dans le fond de kystes. En utilisant l'incidence cranio-caudale, elles sont souvent difficiles à discerner. Tandis que, l'incidence de profil permet de démontrer leurs formes caractéristiques: semi-lunaires, en croissants, curvilignes ou linéaires.

i) Microcalcifications de suture

Elles sont des dépôts calciques sur du matériel de suture. Ces microcalcifications sont typiquement linéaires ou tubulaires et présentent généralement des nœuds.

j) Microcalcifications dystrophiques

Elles mesurent habituellement plus de 0.5 mm de diamètre et sont de formes irrégulières.

Ces microcalcifications sont souvent rencontrées dans un sein irradié ou après un traumatisme mammaire. Notons qu'elles représentent les majorités des cas retrouvés en pathologie mammaire.

k) Microcalcifications amorphes ou indistinctes

Elles sont souvent plus ou moins rondes ou en forme de flocons avec de petites tailles et généralement présentent des contours vagues sans forme spécifique.

l) Microcalcifications fine et polymorphes

Elles sont habituellement mieux visibles que les microcalcifications amorphes. Tandis qu'elles sont irrégulières de taille et de forme variables mesurant généralement moins de 0.5 mm de diamètre.

m) Microcalcifications linéaires et ramifiées

Elles mesurent moins de 0.5 mm d'épaisseur et sont irrégulières et présentent des formes parfois linéaires ou curvilignes généralement discontinues, coudées ou branchées.

Nous présentons sur la figure III.1 les différents types de microcalcifications cités précédemment.

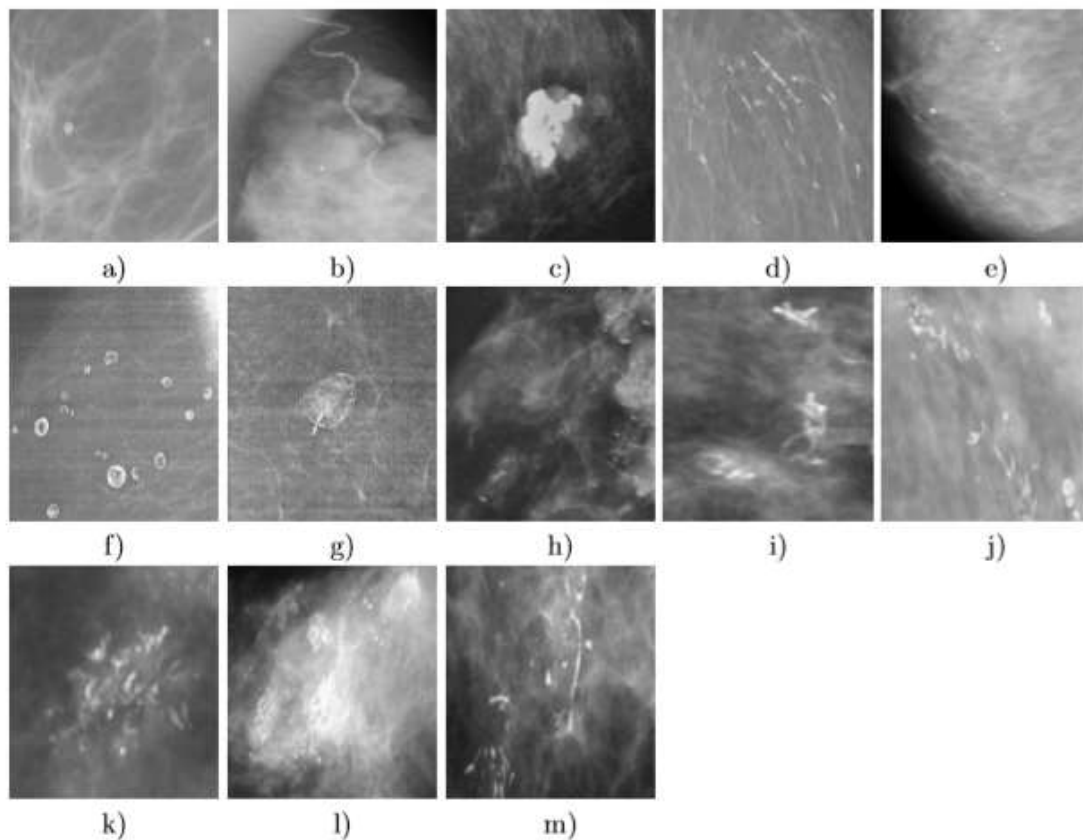


Figure III.1 Les différents types de microcalcification. Source [40]

III.4 Le dépistage du cancer du sein

Le dépistage du cancer du sein consiste en la réalisation des examens de contrôle qui permettent de découvrir des anomalies sans même la présence de symptômes décelables. Cette démarche d'anticipation est très utile puisqu'elle permet de réduire le taux de cas atteints par ce cancer.

Dans ce sens, des campagnes de dépistage du cancer du sein ont vu le jour, proposant d'effectuer des mammographies systématiques et gratuites aux femmes dont l'âge est compris entre 40-50 ans ainsi qu'entre 70-75 ans.

Le Centre International de Recherche sur le Cancer (CIRC) affirme l'efficacité du dépistage mammographie systématique pour la réduction de la mortalité par le cancer du sein [42]. En effet, le taux de réduction a été estimé, chez les femmes dont la tranche d'âge est de 50 à 69 ans choisissant de participer à un programme de dépistage, à 35 %.

Dans les campagnes de dépistage, deux ou quatre images mammaires sont effectuées par patiente, en effet, nous distinguons une ou deux mammographies par sein. L'intérêt majeur est de pouvoir parcourir toute la glande mammaire et d'éviter de rater quelques lésions.

La tumeur ou toute sorte d'anomalie, peut être détectée par le patient lui-même ou par un médecin lors de la consultation durant le processus de dépistage. En effet, les signes d'appel peuvent consister en une masse anormale dans la poitrine et le patient est invité à être inclus dans une expérience clinique continue. [8][35]

L'interrogatoire de la patiente comporte :

- S'informer par rapport aux antécédents familiaux et personnels.
- Connaître le statut ménopausique, prise ou non d'un traitement substitutif.
- Les circonstances de la détection de la tumeur et son évolutivité.

Quant à l'examen clinique mammaire, il inclut notamment :

- Un examen complet des seins (caractéristiques de la tumeur : caractère palpable ou non, taille, topographie, mobilité ainsi qu'un examen du revêtement cutané, du mamelon et de l'aréole), ainsi que des aires ganglionnaires axillaires et sus-claviculaires. [8]
- Recherche de signes d'appel ayant la possibilité d'évoquer une évolution métastatique. [35]

III.5 L'imagerie médicale

Les exigences et les nouveaux défis en dans la santé confèrent à l'imagerie médicale une grande valeur dans les techniques d'investigation en médecine. En effet, l'imagerie médicale présente le procédé avec lequel un médecin peut examiner le dedans du corps humain d'un patient sans l'opérer.

Par ailleurs, l'imagerie médicale est utilisée à des fins cliniques durant l'établissement d'un diagnostic ou durant le traitement de pathologies, voire plus, dans les travaux de recherche scientifique analysant la physiologie des êtres vivants [43].

L'imagerie médicale est sans aucun doute l'un des axes de la médecine qui a le plus progressé ces deux dernières décennies. Eventuellement, le diagnostic s'est nettement évolué suite à l'identification exacte de la lésion facilitant le recours à la chirurgie qui représente la seule solution thérapeutique pour quelques malades.

Si nous donnons une définition de l'imagerie médicale, nous pouvons affirmer que c'est une collection d'images du corps humain acquises à l'aide de différents phénomènes physiques comme la résonance magnétique, les ultrasons et les rayons gamma. En effet, les tissus, les organes, les cellules et l'intégralité de l'anatomie se visualisent sur un ordinateur.

L'imagerie médicale est un outil de diagnostic très pertinent, omniprésent en recherche, effectivement, l'ultime objectif est d'obtenir une image médicale révélant un ensemble d'informations sur un organe tel que le sein. L'image acquise à partir de la texture de cet organe est analysée directement en l'affichant à l'écran [43] [44].

III.6 La mammographie

La mammographie est obtenue suite à l'atténuation du faisceau de rayons X traversant divers tissus mammaires. En effet, les rayons rencontreront ou bien un composant gras saturé de lipides, ou bien un composant gras obscurci sur la plaque du fait de sa densité, ou bien un composant désignant une fibre fibreuse rayonnante constituant la lésion mammalienne. A cet effet, la mammographie apparaît sombre alors que les composants de la lésion (une masse) apparaissent plus prononcés.

Nous présentons sur la figure III.2 l'équipement nécessaire à la réalisation d'une mammographie.



Figure III.2 Equipement de la réalisation d'une mammographie. Source [43]

Tandis que sur la figure III.3, nous présentons le positionnement du sein durant la réalisation d'une mammographie.

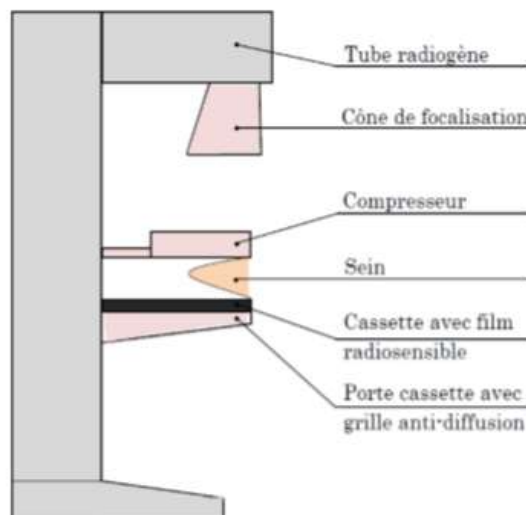


Figure III.3 Positionnement du sein durant la génération d'une mammographie. Source [40]

La mammographie est, de loin, considérée comme l'outil le plus fiable pour le diagnostic du cancer du sein. En effet, elle a un rôle majeur dans la détection précoce de ce dangereux

cancer en montrant les changements opérant dans le sein jusqu'à environ deux ans avant qu'une patiente ou un médecin ne puissent les sentir.

Les signes d'anomalies ont été classés en trois catégories [43] :

- Les densités asymétriques.
- Les masses dont on précise la forme, la taille, les caractéristiques des bords et les distorsions architecturales.
- Les micro-calcifications dont on peut préciser la taille, le nombre, la morphologie, la distribution et l'hétérogénéité.

Il est important de mentionner que, parmi ces trois signes d'anomalies, les de micro-calcifications demeurent les premiers signes d'alerte du cancer du sein.

III.6.1 Les types de mammographies

Il y a trois types de mammographie où chacun est réalisé pour des raisons différentes [6][35] :

- **La mammographie de dépistage**

Elle est réalisée pour les femmes qui ne présentent aucun symptôme de cancer du sein. En effet, les deux seins sont examinés où l'ultime but est de détecter des masses assez petites pour être palpables. La mammographie de dépistage vise à surveiller les changements mammaires au fil des années et facilite la découverte du cancer du sein à un stade précoce.

- **La mammographie diagnostique**

Elle est réalisée pour diagnostiquer une masse ou une zone suspecte. En effet, des images détaillées et des images de la poitrine sont autorisées sous ces angles, pour une analyse précise de la poitrine.

- **La mammographie numérique**

C'est une mammographie électronique plutôt que radiographique.

En effet, elle est réalisée de la même manière que l'imagerie par rayons X, sauf que les images sont stockées sur l'ordinateur.

Nous présentons sur la figure III.4, un schéma désignant le processus de réalisation de la mammographie numérique.

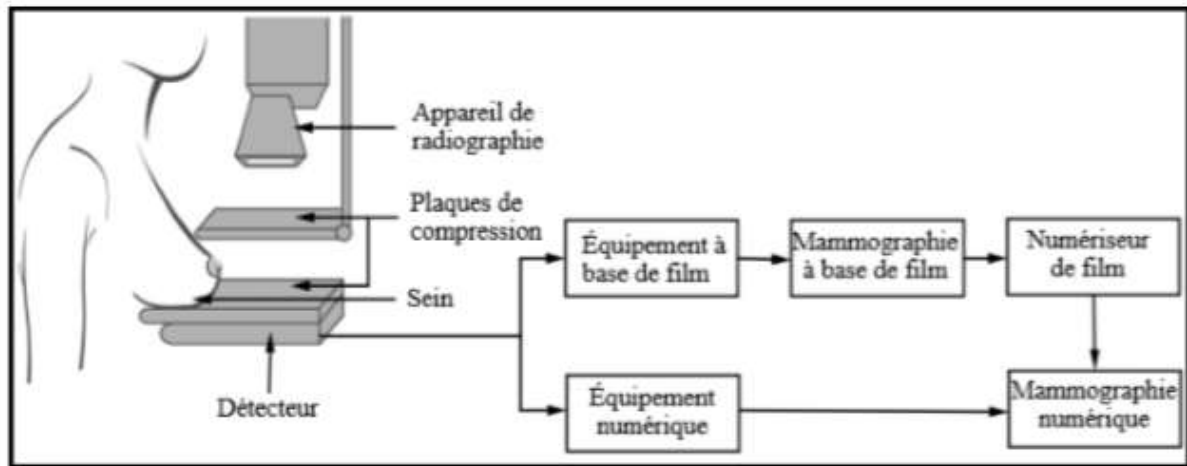


Figure III.4 Processus de réalisation de la mammographie numérique. Source [6]

III.6.2 Les avantages de la mammographie numérique [5]

Nous pouvons citer :

- L'amélioration du contraste.
- La possibilité de réaliser un post traitement à l'image.
- La possibilité de lire, de stocker et de transférer la mammographie.
- La tomo-synthèse et autres technologies de visualisation 3D.
- Baissement de la compression du sein.
- Surtout, l'amélioration de la qualité de l'image.

III.7 Conclusion

Le cancer du sein est l'un des problèmes de santé publique les plus importants menaçant la vie de la majorité des femmes.

Plusieurs pays dans le monde organisent des campagnes de dépistages, en effet, l'ultime but est de détecter les anomalies le plus tôt possible, au moyen de la mammographie, avant qu'elles ne provoquent des symptômes cliniques.

Le développement rapide des technologies d'imagerie médicale a révolutionné la médecine, par conséquent, la vision par ordinateur est intervenue dans le but de permettre aux chercheurs scientifiques et aux médecins d'acquérir des informations vitales suite à l'analyse du corps humain.

L'objectif de notre travail est d'introduire les méthodes de diagnostic assisté par ordinateur (CADx) pour aboutir à un diagnostic plus précis en classifiant des masses en bénignes et malignes dans les cas où la détection et classification est très difficile à l'œil humain.

A cet effet, nous évoquerons dans le prochain chapitre le principe du diagnostic médical du cancer du sein assisté par ordinateur à travers l'analyse numérique de la mammographie pour aider les médecins à établir le diagnostic final.

IV.1 Introduction

L'analyse d'images mammaires réalisée manuellement par un radiologue est très complexe, en effet, ceci est dû principalement au contraste faible qui est l'origine des erreurs de diagnostic effectuées par le radiologue dans 10% à 30% des cas. En outre, certains détails dans la mammographie s'oublient pour le radiologue en raison de la fatigue visuelle qu'il peut avoir.

A cet effet, le développement d'un système d'aide au diagnostic médical assisté par ordinateur est crucial pour la précision du diagnostic final. L'enjeu majeur consiste en l'identification et la classification des masses dans les cas où la détection et la classification se sont avérées compliquées à l'œil humain.

La notion du cancer du sein ainsi que les outils permettant son diagnostic ont été évoqués dans le chapitre précédent. En effet, une telle étude a été accompagnée par la présentation des standards adoptés par les radiologues pour classifier les lésions mammaires en bénignes ou malignes.

Notre objectif dans ce chapitre, est de détailler le principe des systèmes de détection et classification assistées par ordinateur (CADe/CADx) des tumeurs dans la mammographie, voire plus, nous détaillons également l'indexation et la recherche des mammographies par le contenu (CBMIIR).

IV.2 Détection et classification assistées par ordinateur (CADe/CADx) des tumeurs dans la mammographie

Les systèmes de détection et de diagnostic assistés par ordinateur (CAD) des tumeurs dans les mammographies fournissent une seconde interprétation de la mammographie aux radiologistes afin de les aider à détecter / diagnostiquer les lésions malignes à un stade précoce [45]. En effet, l'ultime but est d'offrir une seconde opinion aux radiologues pour augmenter la précision du diagnostic tout en réduisant le temps requis pour l'interprétation des images médicales. Ces systèmes sont classés en deux catégories, à savoir: les systèmes de détection des tumeurs assistée par ordinateur (CADe) et les systèmes de diagnostic ou de classification des tumeurs assistés par ordinateur (CADx).

Les systèmes de détection des tumeurs assistée par ordinateur (CADe) permettent la localisation ou la segmentation des lésions dans les images médicales. De l'autre part, les systèmes de diagnostic des tumeurs assisté par ordinateur (CADx) distinguent la nature des tumeurs : bénignes ou malignes [40].

Il a été démontré que ces systèmes ont un impact positif sur la détection précoce du cancer du sein [41]. En effet, le diagnostic assisté par ordinateur (CAD) surmonte, en particulier, la grande difficulté d'analyse de la texture par les médecins et ainsi il augmente la précision du diagnostic [46].

Différents travaux connexes ont été proposés dans la littérature, où divers systèmes de diagnostic assisté par ordinateur (CAD) ont été développés.

Dans [47], les auteurs proposent un système de diagnostic assisté par ordinateur basé sur une description détaillée de la texture de l'intégralité région de l'image, ceci en commençant des régions externes jusqu'à atteindre les régions internes, au biais de l'extraction de caractéristiques des index phylogénétiques consistant en une branche de la génétique traitant des modifications génétiques au sein des espèces animales ou végétales. Ensuite, la classification est effectuée par l'utilisation de plusieurs classifieurs. Des expériences ont été effectuées pour affirmer les performances de l'approche proposée où une précision de 99,73%, a été atteinte.

Dans [48], les auteurs introduisent un nouveau système de CAD permettant de détecter, segmenter et classifier les masses mammaires au sein des mammographies. En effet, pour la détection, ils ont proposé une cascade de méthodes d'apprentissage en profondeur pour sélectionner des hypothèses affinées en fonction de l'optimisation bayésienne. Par ailleurs, pour la segmentation, ils ont proposé l'utilisation de la méthode de level set. Finalement, pour la classification ils ont utilisé un classifieur d'apprentissage en profondeur. [50][51]

Dans [49], les auteurs affirment que le diagnostic assisté par ordinateur (CAD) aide à obtenir un diagnostic rapide, cohérent et fiable. L'extraction des caractéristiques de Gabor et l'analyse discrète sensible à la localisation (LSDA) ont été appliquées. Lors de la phase de classification, plusieurs classifieurs tels que le k-NN, le classifieur naïve de Bayes (NBC), réseau de neurones probabiliste (PNN) et pleins d'autres ont été appliqués. Suite à l'évaluation des performances, les auteurs constatent que le classifieur k-NN est le plus performant et que le système peut être utilisé dans les hôpitaux et les polycliniques pour aider les cliniciens à vérifier leur diagnostic manuel.

Dans [41], les auteurs proposent des nouveaux systèmes de CAD basés sur des nouvelles méthodologies en mode B ainsi qu'en échographies modernes consistantes en l'élastographie et les ultrasons mammaires automatisés. Des caractéristiques de texture invariantes en niveaux de gris ont été utilisées où la classification des tumeurs malignes était performante.

Une nouvelle technique exploitée dans les systèmes de CAD afin d'améliorer la précision et l'efficacité des programmes de dépistage pour la détection et le diagnostic du cancer du sein est la fractale. L'approche fractale consiste à identifier des structures géométriques présentant une auto-similarité à différentes échelles par la détection des lésions dans les mammographies. En effet, dans le cas des tissus fibroglandulaires, les deux seins sont symétriques et donc pour chaque densité asymétrique, une évaluation de la malignité devient nécessaire. Dans [49], les auteurs définissent de nouvelles caractéristiques fractales basées sur l'extraction d'informations asymétriques à partir de lésions où un classifieur SVM a été appliqué et une grande précision dans la classification des mammographies a été obtenue.

IV.2.1 Détection des tumeurs assistée par ordinateur (CADe)

La détection des tumeurs assistée par ordinateur (CADe) est la première étape majeure dans le diagnostic de la mammographie. Elle consiste à détecter et localiser les lésions suspectes dans le tissu parenchyme mammaire par l'utilisation des descripteurs mathématiques spécifiques.

En effet, les algorithmes les plus habituellement appliqués pour l'extraction des zones d'intérêt (ROI: region of interest) sont basés sur l'analyse des pixels.

Effectivement, les pixels à l'intérieur d'un tissu pathologique présentent des descripteurs différents par rapport aux autres pixels du tissu sein sans la présence d'anomalie, en outre, nous distinguons d'autres approches de segmentation basées régions ainsi que d'autres approches basées contours. Les descripteurs peuvent comprendre: des valeurs de gris, des mesures de tissu ou des mesures morphologiques [7]. Par ailleurs, la classification des zones d'intérêt (ROI) en tissu normal ou tissu pathologique est la phase décisive des CADe.

IV.2.2 Diagnostic des tumeurs assisté par ordinateur (CADx)

Le diagnostic assisté par ordinateur (CADx) désigne la seconde étape dans le diagnostic de la mammographie. Les CADx caractérisent une lésion détectée préalablement par le radiologue ou par les CADe, en donnant comme sortie la classification des lésions en malignes / bénignes ou bien tissu parenchyme mammaire sein.

Les systèmes de diagnostic assisté par ordinateur (CADx) font appel aux approches du traitement d'image, en effet, ils réalisent l'extraction des descripteurs à partir des zones suspectes segmentées pour sélectionner par la suite, les plus discriminants qui seront donnés comme entrée à un classifieur évaluant leur nature. En effet, la classification analyse les descripteurs extraits et les classifie en catégories suite à l'application de classifieurs adéquats comme les k plus proche voisins, les séparateurs à vaste marge (SVM) [46].

Il est utile de mentionner que les CADx diminuent relativement le nombre de recommandations biopsiques établis par les médecins en ce qui concerne les tumeurs bénignes.

Nous présentons sur la figure IV.1 une comparaison entre les systèmes CADe et les systèmes CADx.

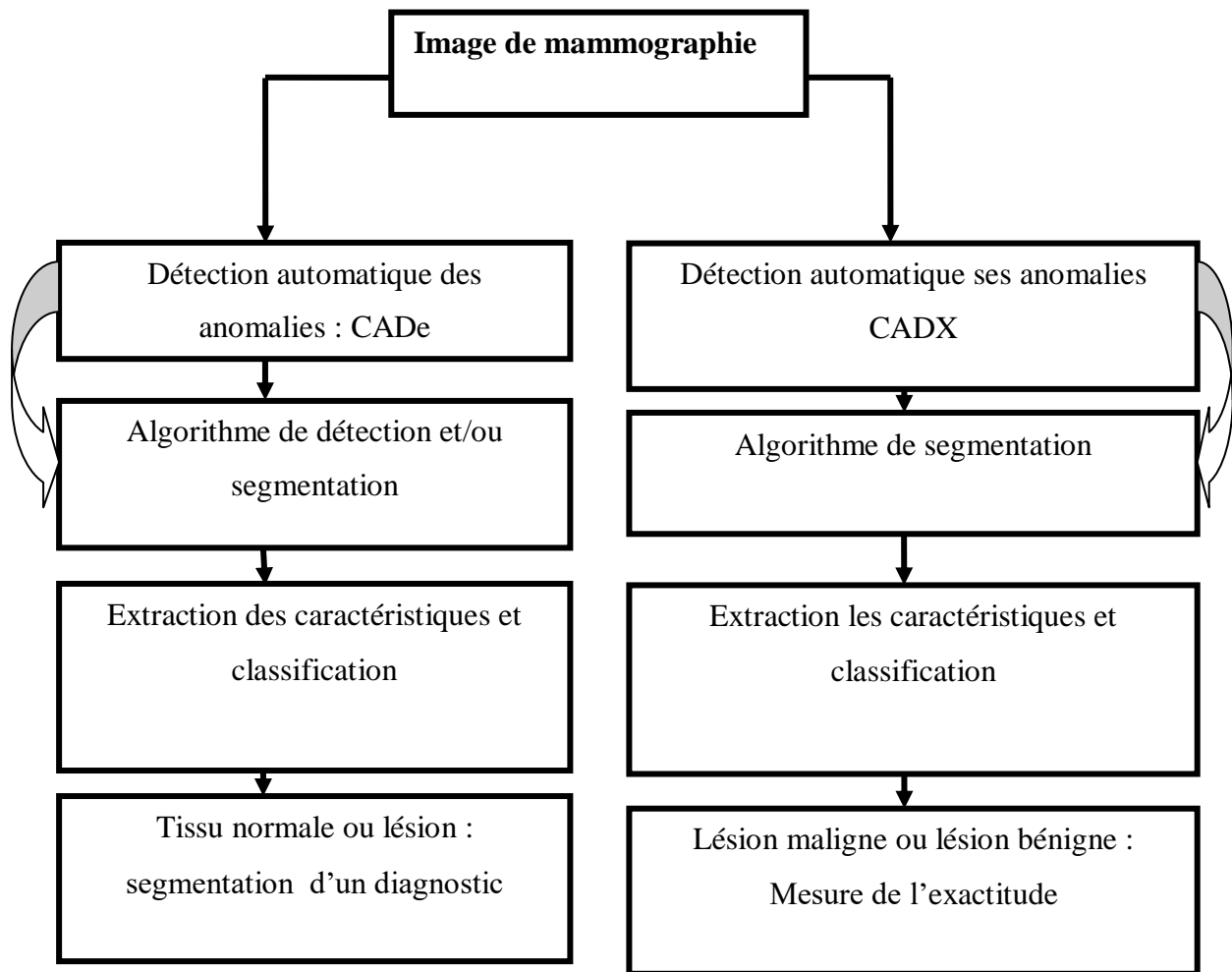


Figure IV.1 La comparaison entre les systèmes CADe et les systèmes CADx. Source [10]

IV.3 Indexation et recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur

De nos jours, vu l'augmentation considérable du nombre des mammographies établies quotidiennement au niveau des établissements sanitaire, la nécessité du développement d'un système d'indexation et de recherche des mammographies par le contenu pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur (CBMIIR : Content Based Mammogram Image Indexing and Retrieval) était en ordre selon le point de vue des chercheurs dans le domaine médical.

L'ultime but des CBMIIR est de rendre pertinent : l'accès, la recherche et la récupération des mammographies archivées. En effet, ils permettent d'accéder efficacement aux mammographies archivées souhaitées, pathologiquement similaires, à la mammographie en cours de diagnostic.

Il est utile de mentionner que les CBMIIR présentent la même architecture des systèmes d'indexation et de recherche des images par le contenu CBIR (Content Based Image Retrieval) [5]. En effet, les images sont stockées avec leurs vecteurs de caractéristiques et lorsqu'une image requête entre dans le module de requête, le vecteur de caractéristiques relatif à cette dernière est extrait. Au cours du processus de recherche, le vecteur des caractéristiques de l'image requête est comparé à l'intégralité des vecteurs de caractéristiques des images stockées dans la base de données. Ensuite, les images les plus similaires par le biais du calcul d'une distance, sont récupérées et renvoyées par le processus de recherche [5]. La figure IV.2 expose le fonctionnement du système d'indexation et de recherche des images par le contenu.

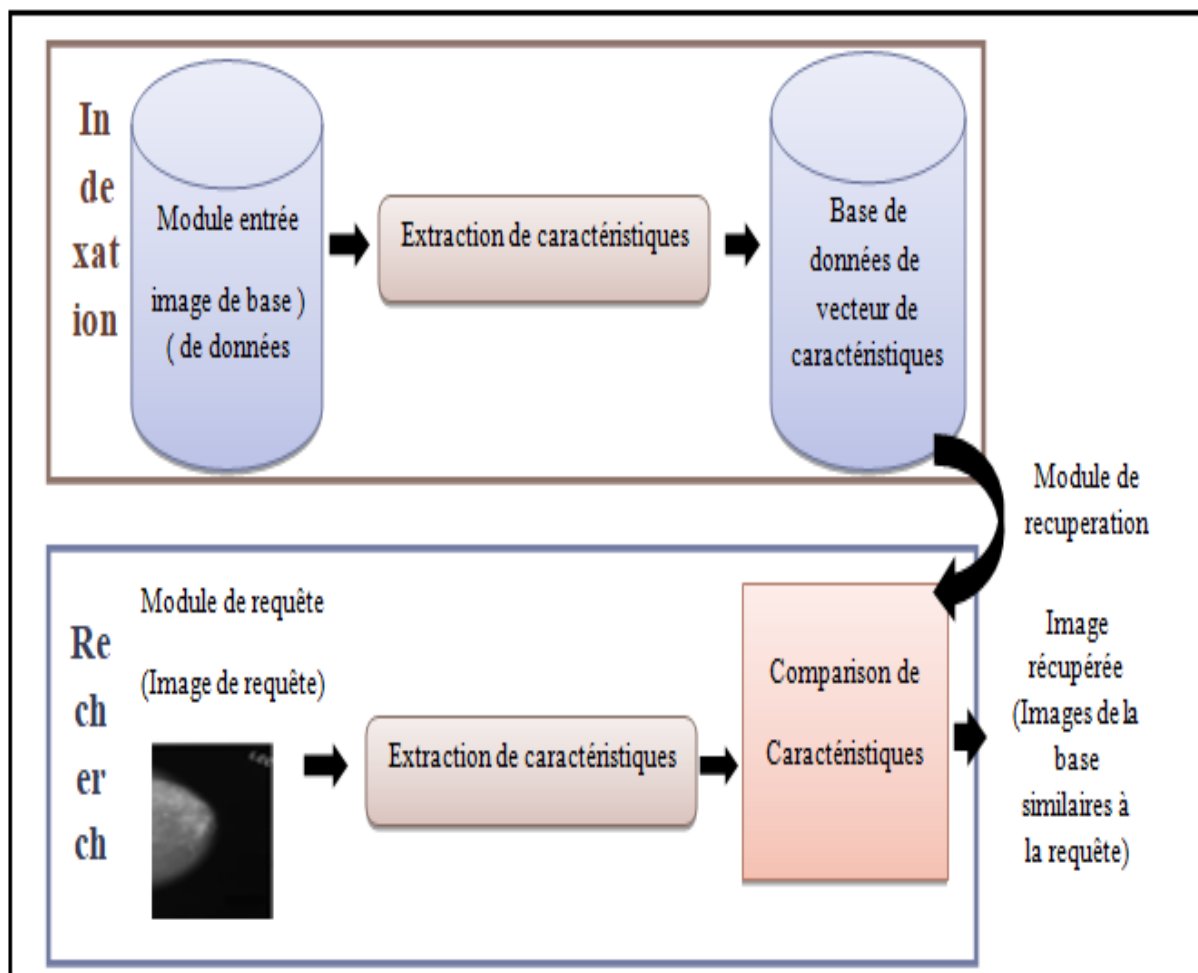


Figure IV.2 Architecture d'un système d'indexation et de recherche des images par le contenu. Source [37]

IV.4 L'apport d'un système d'indexation et de recherche de mammographies par le contenu (CBMIIR) par rapport à un système d'aide au diagnostic assisté par ordinateur (CADx) se basant sur la sortie d'un classifieur

Les systèmes d'aide diagnostic assisté par ordinateur (CADx) permettent d'offrir un deuxième avis aux radiologues durant l'établissement du diagnostic. En effet, ces systèmes réalisent l'extraction des caractéristiques à partir des lésions pour la discrimination entre les masses bénignes, malignes ainsi que le tissu sein.

Il a été démontré que les systèmes CADx génèrent un taux relativement élevé de faux positifs ainsi que de faux négatifs [5] ce qui mène les radiologues à accroître le nombre de biopsies inutiles durant la prise de décision.

Une alternative a été proposé dans ce sens consistant en les systèmes d'indexation et recherche des images mammaires par le contenu (CBMIIR : Content Based Mammogram Image Indexing and Retrieval).

En effet, ces systèmes offrent une décision basée sur la comparaison visuelle où le diagnostic se fait à base de cas : durant le processus de recherche, nous récupérons les images mammaires archivées, étant les plus similaires à l'image mammaire requête.

Les radiologues durant leur formation acquièrent l'expérience de distinction entre les pathologies mammaires par la lecture d'énorme nombre de mammographies, en effet, ils se sentent plus familiarisés avec le diagnostic à base de cas similaires.

Si nous allons analyser la différence entre un système CADx et un système CBMIIR, nous affirmons que la seule différence réside dans la présentation du diagnostic. En effet, l'enjeu majeur vise à renforcer la performance du diagnostic assisté par ordinateur (CADx) qui se base sur un résultat abstrait numérique généré par un classifieur et dont les radiologues trouvent une ambiguïté à interpréter et surtout à en être convaincu [5].

Nous allons approfondir ce point : dans les systèmes d'aide au diagnostic médical (CADx), un classifieur va être optimisé sur des données d'apprentissage, ensuite; la classification d'un nouveau cas aura lieu. Donc, les radiologues reçoivent l'aide sous forme de symboles

numériques abstraits faisant référence par exemple à la classe où appartient la région d'intérêt requête.

D'autre part, les radiologues se sentent plus à l'aise et confiants lorsqu'ils établissent leur diagnostic final en se basant sur des cas déjà diagnostiqués et similaires, au cas en cours de diagnostic. En effet, ces derniers sont récupérés et affichés à partir de bases de données de référence simultanément avec le résultat de la classe à qui appartient le cas en cours d'analyse. De plus, les radiologues pourront indexer les nouveaux cas diagnostiqués. Dans [5], un état de l'art des approches d'indexation et de recherche des mammographies par le contenu (CBMIIR) est établi pour approfondir les connaissances dans ce sens.

IV.5 Conclusion

En raison du dépistage non précis du cancer du sein ainsi que la difficulté de procéder à une double lecture dans certaines lésions où les radiologistes trouvent de l'ambiguïté, il est nécessaire que le CAD prenne une place importante vu les diagnostics très prometteurs qu'il offre.

Dans le chapitre suivant, nous allons détailler l'approche contributive pour la caractérisation de texture des zones d'intérêt extraites des mammographies ainsi que pour le développement de notre système d'aide au diagnostic médical du cancer du sein assisté par ordinateur.

V.1 Introduction

De nos jours, les progrès scientifiques dédiés au traitement d'images et à la vision par ordinateur ont permis le diagnostic du cancer du sein assisté par ordinateur. En effet, l'indexation et la recherche d'images médicales par le contenu offrent aux radiologues la possibilité d'accéder, d'interroger et de manipuler directement les bases de données d'images médicales archivées à l'aide de leur contenu.

Dans ce chapitre, nous évoquerons le développement du système «Mammo-Zipf-Har» de classification des tumeurs mammaires basées sur la technique Template-matching ainsi que la caractérisation de la texture des images mammaires par une fusion des lois de puissance : Zipf, Zipf inverse et la matrice de co-occurrence d'Haralick.

V.2 Problématique

Notre problématique de recherche se base sur un ultime but de la contribution d'une technique robuste d'analyse de la texture d'images mammaires. En effet, les relations linéaires sont les plus simples à appliquer dans tous les domaines de recherche. Néanmoins, l'inconvénient majeure de ces relations est leur limites durant la modélisation d'une image vu sa structure assez complexe.

En comparant les modèles linéaires comme par exemple un polynôme avec les modèles non linéaires, nous affirmons que ces derniers sont rarement exploités comme par exemple : les lois gaussiennes, les lois exponentielles et les lois puissance.

Les lois de puissance Zipf et Zipf inverse ont été appliquées dans un processus d'aide au diagnostic médical du cancer du sein assisté par ordinateur pour la caractérisation de la texture des images mammaires. En effet, ces lois ont permis de générer des descripteurs texturaux discriminants pour la distinction entre diverses textures mammaires : tissu portant une tumeur maligne, tissu portant une tumeur bénigne et un tissu ne possédant aucune pathologie [5].

Nous avons tiré une variantes des perspectives de ces travaux en proposant une fusion des lois de Zipf et de Zipf inverse avec l'approche de la matrice de co-occurrence d'Haralick pour bénéficier de l'apport complémentaire de ces deux approches de caractérisation de la texture dans le but d'aboutir à une caractérisation pertinente des textures mammaires.

Vu le nombre important des mammographies réalisées quotidiennement par les radiologues, l'exploitation efficace de ces ressources demeure un défi où la conception des systèmes qui permettent une indexation et une recherche efficace de ces images, semble la solution adéquate. En effet, les radiologues ont manifesté leur insatisfaction du résultat abstrait généré par un classifieur dans les systèmes d'aide au diagnostic médical du cancer du sein (CADx). A cet effet, nous visons à renforcer les systèmes (CADx) par la nouvelle tendance des systèmes d'indexation et de recherche des mammographies par le contenu (CBMIIR) dont les radiologues expriment une satisfaction vu que le diagnostic sera effectué à base de cas récupérés et affichés à partir de bases de données indexées de cas préalablement diagnostiqués.

L'objectif visé est le développement d'un système d'indexation et de recherche des images mammaires par le contenu basé sur la fusion des lois de Zipf et Zipf inverse ainsi que les matrices de co-occurrence d'Haralick pour caractériser les textures des zones d'intérêt

extraites à partir des mammographies pour aboutir à un diagnostic assisté par ordinateur basé sur la technique Template-matching.

V.3 Approche proposée basée fusion des lois de puissance : Zipf, Zipf inverse et la matrice de co-occurrence d'Haralick pour la classification des zones d'intérêt des mammographies

Dans le domaine du traitement d'image, la description d'une image consiste à extraire les descripteurs qui décrivent au mieux et de façon quantitative ou qualitative les objets présents dans l'image. Dans notre approche nous allons fusionner les descripteurs texturaux générés par les lois de Zipf et de Zipf inverse avec les descripteurs développés par Haralick dans sa méthode de dépendance spatiale des niveaux de gris: SGLD (Spatial Gray Level Dependency) pour tenter d'aboutir à un résultat de classification bien précis [55].

V.3.1 Analyse et caractérisation de la texture des zones d'intérêt par les lois de Zipf et de Zipf inverse

Les lois de Zipf et de Zipf inverse permettent d'analyser et caractériser la complexité structurelle de la texture et de quantifier la structure sous-jacente du contenu des zones d'intérêts segmentées. Elles sont basées sur une approche statistique [6]. Par la suite, un prétraitement réalisé avec le codage des rangs généraux (mentionné dans la section II.5) est appliqué sur les pixels des zones d'intérêt pour obtenir les courbes de Zipf et Zipf inverse, exposées sur la figure V.1. Où cette dernière présente un exemple de courbes de Zipf et de Zipf inverse obtenues à partir de trois zones d'intérêts : l'une sans lésion, une autre portant une tumeur bénigne et la dernière portant une tumeur maligne.

Il est utile de mentionner que pour les images mammaires seins nous avons utilisé un découpage d'une partie du tissu sein.

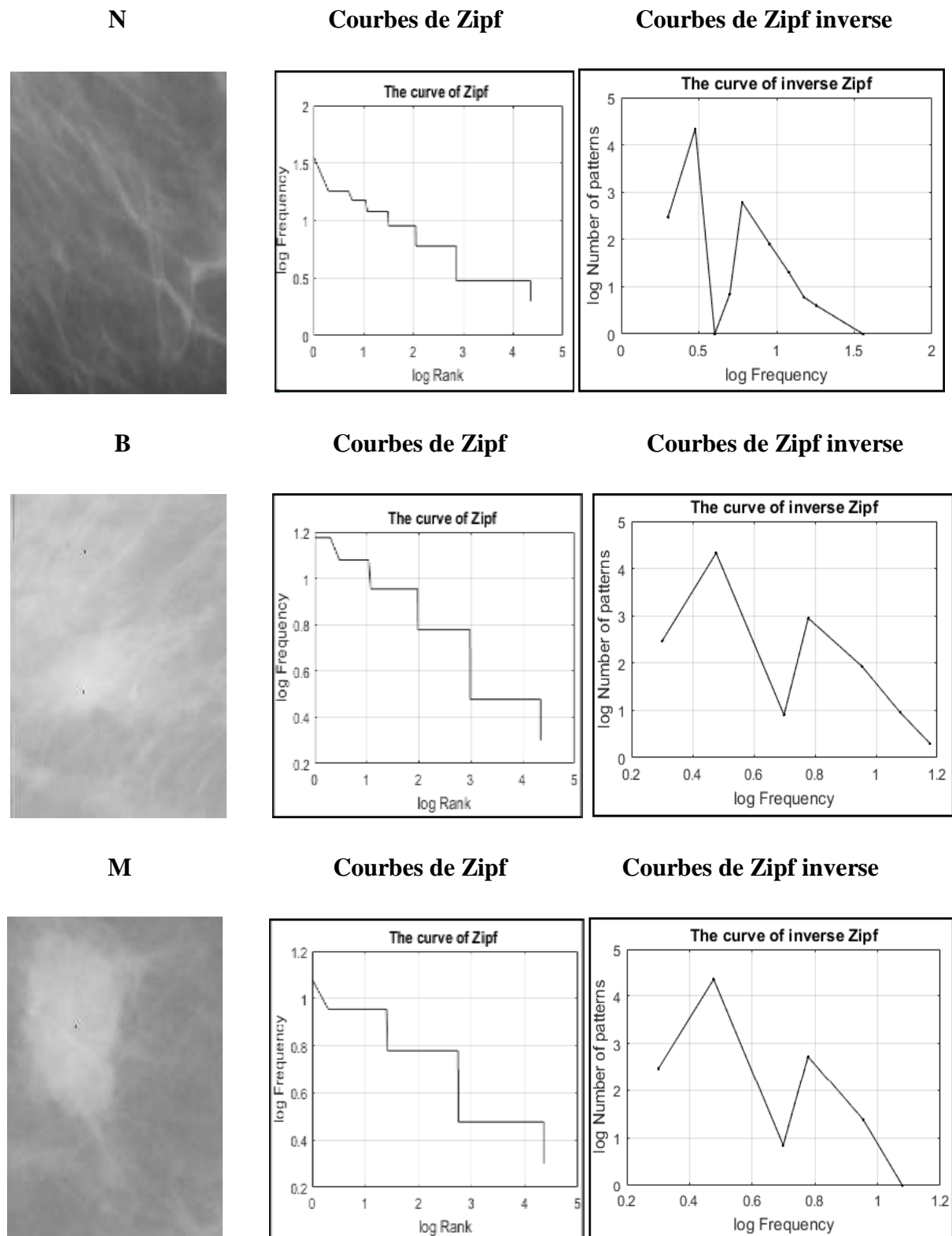


Figure V.1 Courbes de Zipf et de Zipf inverse de trois zones d'intérêts encodées par le codage des rangs généraux : **N** zone d'intérêt sans tumeur, **B** zone d'intérêt portant une tumeur bénigne, et **M** zone d'intérêt portant une tumeur maligne. Source [6].

- **Analyse des courbes de Zipf relatives aux diverses zones d'intérêt**

Dans le cas d'une zone d'intérêt avec le tissu mammaire sein, nous obtenons une ordonnée à l'origine très élevée du fait que le tissu mammaire est homogène et nous distinguons ainsi un motif homogène répété plusieurs fois.

Tandis que pour le cas de la zone d'intérêt portant une tumeur bénigne, nous notons une ordonnée à l'origine inférieure. Par ailleurs, pour la zone d'intérêt présentant une tumeur maligne, nous observons une ordonnée à l'origine de la courbe de Zipf nettement plus basse que celles de la tumeur bénigne du fait que la tumeur maligne présente une texture hétérogène causée par la nature invasive des tumeurs malignes et par conséquent nous ne distinguons pas de sur-représentation du motif homogène, au contraire, plusieurs motifs hétérogènes provoquent que l'ordonnée à l'origine soit basse.

- **Analyse des courbes de Zipf inverse relatives aux diverses zones d'intérêt**

Nous pouvons affirmer que la distinction entre les courbes de Zipf inverses pour la zone d'intérêt avec un tissu normal, la zone d'intérêt avec une tumeur bénigne et la zone d'intérêt avec une tumeur maligne est claire. L'ordonnée à l'origine de la courbe de Zipf inverse de la tumeur bénigne est inférieure à celle de la courbe Zipf inverse de la tumeur maligne, en effet, la région d'intérêt de la tumeur maligne présente une texture complexe, donc le nombre de l'occurrence est très élevée, ce qui fait que la courbe de Zipf inverse comporte un grand nombre de motifs qui apparaissent une fois [6].

Suite à ce qui a été analysé ci-dessus, nous pouvons en déduire que l'analyse des zones d'intérêt par les lois de Zipf et de Zipf inverse à l'aide d'un prétraitement basé sur le codage des images mammaires à base de rangs, permet de distinguer de manière significative les tumeurs malignes ou bénignes ainsi que les tissus seins à l'aide des descripteurs calculés et extraits à partir des courbes de Zipf et de Zipf inverses pour chaque zone d'intérêt.

Les descripteurs dérivés des courbes de Zipf et de Zipf inverses sont les suivants [6] :

- **Les pentes des courbes de Zipf et de Zipf inverse**

La pente moyenne d'une courbe est le coefficient directeur de la droite des moindres carrés. Elle est donnée par la formule V.1 :

$$p = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (\text{V.1})$$

- **L'aire délimitée par la courbe de Zipf**

Nous calculons l'aire délimitée par la courbe de Zipf à partir des courbes de Zipf obtenues suite au codage de l'image par le codage des rangs généraux. Soit n le nombre de motifs de la courbe, f_i la fréquence et r_i le rang du motif i , l'aire de la courbe est donnée par la formule V.2 :

$$A = \sum_{i=1}^{n-1} \frac{(f_i + f_{i+1})(r_{i+1} - r_i)}{2} \quad (\text{V.2})$$

- **Entropie1 de la courbe de Zipf**

L'entropie relative aux motifs des images mammographies est définie par la formule V.3:

$$H_w = - \sum_{r=1}^R \frac{f(r)}{T} \log_R \frac{f(r)}{T} \quad (\text{V.3})$$

Dans cette formule, $f(r)$ représente la fréquence du motif pour la ligne r , T représente le nombre total de motifs différents, et nous utilisons un logarithme avec la base R .

- **Entropie2 de la courbe de Zipf**

L'entropie relative à la fréquence d'apparition des motifs est définie par la formule suivante:

$$H_f = - \sum_{f=1}^F \frac{I(f)}{R} \log_F \frac{I(f)}{R} \quad (\text{V.4})$$

Dans cette formule, $I(f)$ représente le nombre de motifs distincts ayant une fréquence d'apparition égale à f et F représente le nombre entier d'occurrences des motifs dans l'image.

- Les ordonnées à l'origine des courbes de Zipf et Zipf inverse
- Le constant alpha de la courbe de Zipf

La loi de Zipf est fortement exprimée de la façon suivante: Quel que soit un motif appartenant à une image, la fréquence d'apparition de ce motif * son rang dans une liste ordonnée décroissante des fréquences d'apparition des motifs = constante [6].

V.3.2 Analyse et caractérisation de la texture des zones d'intérêt par la matrice de co-occurrence d'Haralick

En raison de leur richesse en informations de texture, les matrices de co-occurrences demeurent les plus connues et les plus utilisées pour l'analyse de la texture [55][19].

Elles estiment des propriétés des images relatives à des statistiques de second ordre.

Une matrice de co-occurrence mesure la probabilité d'apparition des paires de valeurs de pixels situés à une certaine distance dans l'image comme mentionné sur la figure V.2. Elle est basée sur le calcul de la probabilité $P(i; j; \theta; \delta)$ qui représente le nombre de fois où un pixel de niveau de couleur i apparaît à une distance relative δ d'un pixel de niveau de couleur j et selon une orientation θ donnée.

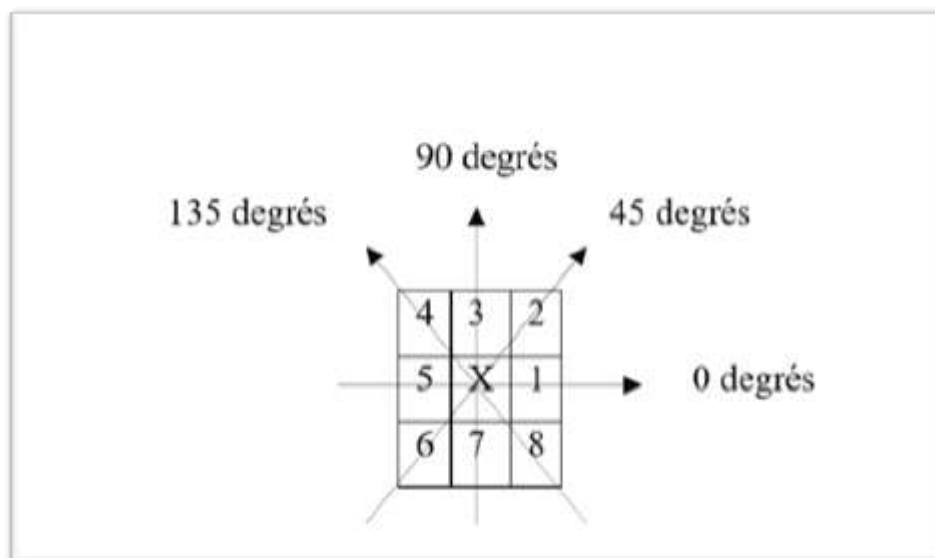


Figure V.2 Plus proches voisins du pixel 'x' selon 4 directions. Source [19].

Par exemple, les plus proches voisins de 'x' selon la direction $\theta=135$ degrés sont les pixels 4 et 8.

Chapitre V : Conception et Implémentation

Les caractéristiques extraites à partir de ces matrices contiennent des informations notamment sur l'homogénéité, les dépendances linéaires entre les niveaux de gris, le contraste et la complexité de cette image.

Les matrices obtenues selon ces quatre directions sont alors calculées comme dans (1), (2), (3) et (4) où (k, l) sont les coordonnées d'un pixel de niveau de couleur $i \in [0, n^{\max} - 1]$ et (m, n) celles du pixel de niveau de couleur $j \in [0, n^{\max} - 1]$

Les figure V.3 et V.4 montrent un exemple de calcul des $P(i, j)$ à partir d'une petite image de dimension 4×4 composée de quatre niveaux de gris (0, 1, 2, 3). Cet exemple se limite au cas $\delta = 1$ et $\theta = 0$. L'élément $(2, 3)$ de la matrice $P(1, 0)$ est égal à 4, cela signifie qu'il existe 4 configurations dans l'image où un pixel de niveau de gris 2 est séparé horizontalement d'un autre pixel de niveau de gris 3 par une distance 1. Ces configurations sont représentées en trait gris sur l'image.

$$P(i, j, \delta, 0) = \left| \left\{ ((k, l), (m, n)) \in (N \times M)^2 \mid (k - m = 0, |l - n| = \delta, I_{k,l} = i, I_{m,n} = j) \right\} \right| \quad (1)$$

$$P(i, j, \delta, 45) = \left| \left\{ ((k, l), (m, n)) \in (N \times M)^2 \mid \left((k - m = \delta, l - n = -\delta) \vee (k - m = -\delta, l - n = \delta), I_{k,l} = i, I_{m,n} = j \right) \right\} \right| \quad (2)$$

$$P(i, j, \delta, 90) = \left| \left\{ ((k, l), (m, n)) \in (N \times M)^2 \mid (|k - m| = \delta, l - n = 0, I_{k,l} = i, I_{m,n} = j) \right\} \right| \quad (3)$$

$$P(i, j, \delta, 135) = \left| \left\{ ((k, l), (m, n)) \in (N \times M)^2 \mid \left((k - m = \delta, l - n = \delta) \vee (k - m = -\delta, l - n = -\delta), I_{k,l} = i, I_{m,n} = j \right) \right\} \right| \quad (4)$$

Figure V.3 Les quatre directions pour obtenir la matrice co-occurrence. Source [19]

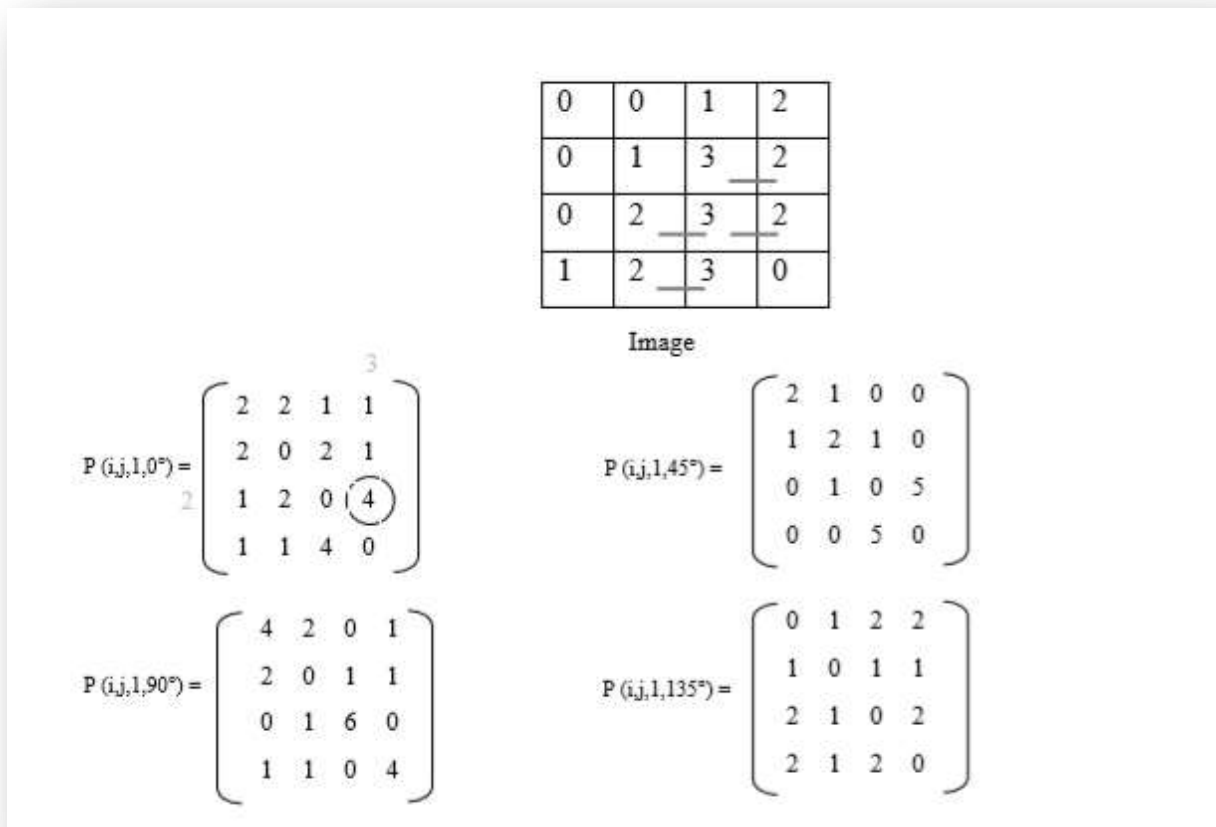


Figure V.4 Exemple de matrices de co-occurrence construites à partir d'une image 4×4 composée de 4 niveaux de gris. Source [19]

La plupart des images sont codées sur 256 niveaux de gris, par conséquent, la taille des matrices de co-occurrence est de 256×256. Nous pouvons affirmer que ces matrices comptabilisent une très grosse quantité d'informations difficile à l'exploit directe. Ainsi, nous extrayons de l'information à partir de ces matrices afin de mettre en évidence la structure des textures. Dans [39], la proposition de 14 paramètres, caractérisant les textures, issus de ces matrices a été effectuée.

Nous allons évoquer les 6 paramètres considérés comme étant les plus utilisés et les plus pertinents [19] :

- **L'énergie**

$$ENE = \sum_i \sum_j (P_{ij}(\delta, \theta)^2)$$

(V.5)

Ce paramètre mesure l'uniformité de la texture. Il atteint de fortes valeurs lorsque la distribution des niveaux de gris est constante ou de forme périodique. Dans ce dernier cas, les valeurs élevées d'énergie sont obtenues pour les matrices $P(d, \theta)$ lorsque (d, θ) correspond à la période.

- **Le contraste**

La valeur en est d'autant plus élevée que la texture présente un fort contraste. Ce paramètre est fortement non corrélé à l'énergie.

$$CST = \sum_i \sum_j ((i - j)^2 P_{ij}(\delta, \theta))$$

(V.6)

- **L'entropie**

Ce paramètre mesure le désordre dans l'image. Contrairement à l'énergie, l'entropie atteint de fortes valeurs lorsque la texture est complètement aléatoire (sans structure apparente). Elle est fortement corrélée (par l'inverse) à l'énergie.

$$ENT = -\sum_i \sum_j (\log P_{ij}(\delta, \theta) P_{ij}(\delta, \theta))$$

(V.7)

- **La variance**

La variance mesure l'hétérogénéité de la texture. Elle augmente lorsque les niveaux de gris diffèrent de leur moyenne. La variance est indépendante du contraste.

$$VAR = \sum_i \sum_j ((i - \mu)^2 P_{ij}(\delta, \theta))$$

(V.8)

- **La corrélation**

La corrélation mesure la dépendance linéaire (relativement à (δ, θ)) des niveaux de gris de l'image.

La corrélation n'est corrélée ni à l'énergie, ni à l'entropie.

$$COR = \sum_i \sum_j \left(\frac{(i - \mu)(j - \mu) P_{ij}(\delta, \theta)}{\sigma^2} \right)$$

(V.9)

- **Le moment inverse**

$$IDM = \sum_i \sum_j \frac{P_{ij}(\delta, \theta)}{1 + (i - j)^2}$$

(V.10)

IDM (Inverse Difference Moment) mesure l'homogénéité de l'image. Ce paramètre est corrélé à une combinaison linéaire des variables ENE et CST.

La méthode d'extraction de ces paramètres basée sur le calcul des matrices de co-occurrence est une des méthodes les plus proches de la notion de texture. Elles mettent effectivement en avant les relations qui existent entre les pixels de l'image en faisant intervenir l'aspect local (les niveaux de gris) et l'aspect spatial $((\delta, \theta))$.

V.3.3 Comparaison entre les lois de Zipf et Zipf inverse et la matrice de co-occurrence pour l'analyse de la texture d'image

Nous présentons sur le tableau V.1, une comparaison entre les lois de Zipf et de Zipf inverse ainsi que la matrice de co-occurrence d'Haralick.

Loi de Zipf et Zipf inverse	Matrice Co-occurrence
Reposent sur un approche statistique, elles permettent de qualifier la structure sous-jacente du contenu d'une image, les paramètres extrait de la loi estimée traduisent la texture au sein d'une image.	Fait ressortir la fréquence et la régularité d'une teinte dans une image obligatoirement en niveau de gris, ce qui permet de détecter un éventuel motif et par conséquent une texture.
Utilisent un calcul de certains indices permettant de caractériser la texture au sein des images.	Utilise un calcul de certains indices permettant de caractériser la texture au sein des images.
Illustrent leur comportement par une représentation graphique dite courbe de Zipf ou de Zipf inverse, à partir desquelles l'extraction des paramètres est réalisée pour la représentation des images.	Les paramètres de la matrice de co-occurrence sont calculés à partir de cette dernière qui sont reconnus discriminant.
Application récente dans le domaine du traitement d'image (2002, Yves Caron)	Elle est l'une des méthodes les plus populaires en analyse de texture (Haralick et al., 1973).
Permettent de déterminer la fréquence d'apparition d'un motif formé des 9 pixels (le motif qui a donné les meilleurs résultats) selon un balayage séquentiel des niveaux de gris d'une image.	Permet de déterminer la fréquence d'apparition d'un motif formé de 2 pixels séparés par une certaine distance d dans une direction particulière par rapport à l'horizontal.
Effectuent leur calcul directement sur l'image après l'avoir codé par un codage pour réduire le nombre des niveaux de gris	L'image doit être normalisée avant le calcul des différents paramètres de la texture.

susceptible d'être présent dans l'image pour tracer les courbes.	
Zipf et Zipf inverse sont des outils utilisés pour effectuer une analyse statistique d'une image mammaires.	La matrice de co-occurrence est un outil utilisé pour effectuer une analyse statistique d'une image.
Les courbes de Zipf permettent leur exploitation directe pour le calcul des descripteurs.	La matrice de co-occurrence contient une masse très importantes d'information difficilement manipulable, c'est pour cela qu'elle n'est pas utilisée directement mais à travers des mesures dites indices de textures.
La pertinence des lois de Zipf est basée sur le choix de la taille du motif et du codage appliqué à ce dernier.	La réussite de la méthode de la matrice de co-occurrence est basée sur le choix du vecteur de déplacement et de la taille de la fenêtre de voisinage.

Tableau V.1 Illustration d'une comparaison entre l'approche d'Haralick et celle des Zipf et Zipf inverse pour l'analyse de la texture d'image.

V.4 La classification

La classification est généralement la dernière étape du système d'aide au diagnostic assisté par ordinateur (CADx) ou d'un système d'indexation et de recherche des mammographies par le contenu (CBMIIR), qui utilise les caractéristiques extraites de l'image dans la classification automatique des objets d'image.

Le processus de classification sert à étiqueter les échantillons de la base de données par un certain nombre de descripteurs discriminants afin de l'identifier. Les échantillons peuvent être des pixels ou un espace dans l'image ou l'objet représenté dans l'image ou dans l'image entière. La classification a deux buts qui sont [40] :

- La classification des pixels d'images en plusieurs segments d'objets .Citons l'exemple de classification d'images mammaire en lésions ou sans lésions
- La classification d'image selon différentes catégories. Citons l'exemple de classification des zones d'intérêt mammaires en malignes ou bénignes.

Nous pouvons distinguer deux catégories de méthodes de classification: les classification non-supervisées et celles supervisées [40].

V.4.1 Les méthodes de classification non-supervisées

Les méthodes de classification non-supervisées nommées automatiques, ou groupement connue en anglais par clustering consistent à attribuer les images aux différentes classes naturellement sans aucune connaissance préalable en identifiant une structure dans les images en se basant sur leurs contenus selon deux critères essentiels qui sont la grande homogénéité de chaque classe et la bonne séparation entre les classes. Les méthodes les plus communément utilisées : k-moyennes, le regroupement hiérarchique [40] [53].

V.4.2 Les méthodes de classification supervisées

Ces techniques sont utilisées quand il y'a une connaissance préalable sur les images à étudier (tel est le cas des images mammaires). La classification supervisée est effectuée sur un groupe d'individus dont chaque groupe appartient à une classe connue. Ces individus constituent des échantillons "d'apprentissage". Ils sont utilisés pour entraîner le classifieur. D'autres échantillons, dits "de test", servent à valider la classification en évaluant sa pertinence à travers le taux d'individus bien classés. Il existe plusieurs méthodes de classification supervisées. Les méthodes les plus réputées: Machine à Vecteurs de Support (SVM), Réseau de Neurones, Méthode des K Plus Proches Voisins (KPPV), Arbre de Décision.

Pour élaborer notre travail, nous utilisons les méthodes de classification supervisées basée Template-matching à travers le classifieur des K Plus Proches Voisins (KPPV) ainsi que les Machines à Vecteurs de Support (SVM).

- **Les machines à vecteurs de support (SVM)**

Appelées aussi les séparateurs à vaste marge .Le principe de l'algorithme SVM est basée sur la séparation des points appartenant à différentes classes, par un hyperplan optimal qui assure la distance maximale de points de données. Les points qui ne sont pas linéairement

séparables, sont projetés dans un nouvel espace en appliquant une fonction du noyau (kernel). [19].

Les fonctions et l'espace du noyau permettent le calcul du produit interne sans projection explicite des vecteurs dans un espace de dimension supérieure. Il existe plusieurs méthodes pour faire la classification multiclassées [61] [62].

V.5 Etapes suivies et accomplies pour la réalisation du système Mammo-Zipf-Har d'indexation et la recherche des mammographies par le contenu

L'ultime but de notre système d'indexation et de recherche des mammographies basé sur le contenu Mammo-Zipf-Har est la prise des décisions lors de la récupération d'images similaires, au cas en cours de diagnostic, indexées dans une base de données de référence.

V.5.1 Caractérisation de la texture par les lois puissance Zipf et Zipf inverse et la matrice de co-occurrence d'Haralick

Nous allons utiliser simultanément les deux méthodes d'analyse de la texture en fusionnant les descripteurs obtenus suite à l'analyse des zones d'intérêt par les lois puissance Zipf et Zipf inverse ainsi que ceux obtenus à partir de la matrice de co-occurrence d'Haralick, afin de bénéficier de l'apport complémentaire de ces deux méthodes d'analyse de la texture.

V.5.2 Sélection des descripteurs les plus discriminants par les algorithmes génétiques

Nous allons appliquer les algorithmes génétiques (AG) sur les vecteurs descripteurs des zones d'intérêt (8 descripteurs générés par les lois de Zipf et de Zipf inverse et 13 autres générés par la matrice de co-occurrence d'Haralick) pour optimiser la taille du vecteur descripteur en conservant que l'information pertinente et l'obtention de 9 descripteurs non corrélés, par la suite, nous utiliserons ces descripteurs dans les méthodes de modélisation, plus précisément, l'analyse discriminante afin de classifier une zone d'intérêt requête en tumeur maligne ou bénigne ou bien une région saine ne présentant pas de lésion.

Les algorithmes génétiques (Genetic Algorithms) sont des stratégies d'adaptation et des techniques d'optimisation globale [56]. Ce sont des algorithmes évolutionnaires qui s'inspirent des mécanismes d'évolution darwinienne des populations biologiques.

Les algorithmes génétiques ont trois applications majeures nommées la recherche intelligente, l'optimisation et l'apprentissage automatique. Ils travaillent avec un ensemble de solutions

candidates appelées population basés sur le principe darwinien de la «survie du plus apte». Les algorithmes génétiques obtiennent la solution optimale après une série de calculs itératifs.

- **Principe des algorithmes génétiques**

Un AG a généralement quatre composants : une population d'individus, où chaque individu de la population représente une solution candidate au problème d'optimisation, une fonction de fitness qui est une fonction d'évaluation par laquelle nous pouvons dire si un individu est une bonne solution ou non, une fonction de sélection qui décide comment choisir les bons individus de la population actuelle pour la création de la génération suivante et des opérateurs génétiques tels que le croisement et la mutation, qui explorent de nouvelles régions de l'espace de recherche tout en gardant une partie de l'information actuelle.

En général, le fonctionnement d'un AG est basé sur les phases suivantes (Figure V.5)

1. Initialisation : génération aléatoirement d'une population initiale de taille N chromosomes. Nous entrons les descripteurs des images comme initialisations et le meilleur chromosome a le signe 1 si il valide et 0 pour le chromosome non valide pour choisir le best.

2. Evaluation : évaluation chaque individu de la population par la fonction d'évaluation appropriée au problème (fonction de fitness). (Lorsque choisie le chromosome valide).

3. Reproduction : création d'une nouvelle population de N chromosomes par l'utilisation d'une méthode de sélection appropriée et l'application d'opérateurs génétiques (croisement et mutation) sur certains chromosomes au sein de la population courante. Nous choisissons 300 pour la population pour créer beaucoup nouvelles générations.

4. Retour à la phase 2 tant que la condition d'arrêt du problème n'est pas satisfaite.

Le critère d'arrêt a pour but de nous assurer une bonne qualité de la solution finale obtenue par l'algorithme génétique et associée à l'individu de meilleure qualité dans la dernière génération. Le critère d'arrêt peut prendre la forme suivante [58]

- ✓ Le nombre limite de générations autorisées a été atteint.
- ✓ Une stabilité de la population a été atteinte (la population cesse d'évoluer ou n'évolue plus suffisamment) de point de vue de la fonction de fitness.

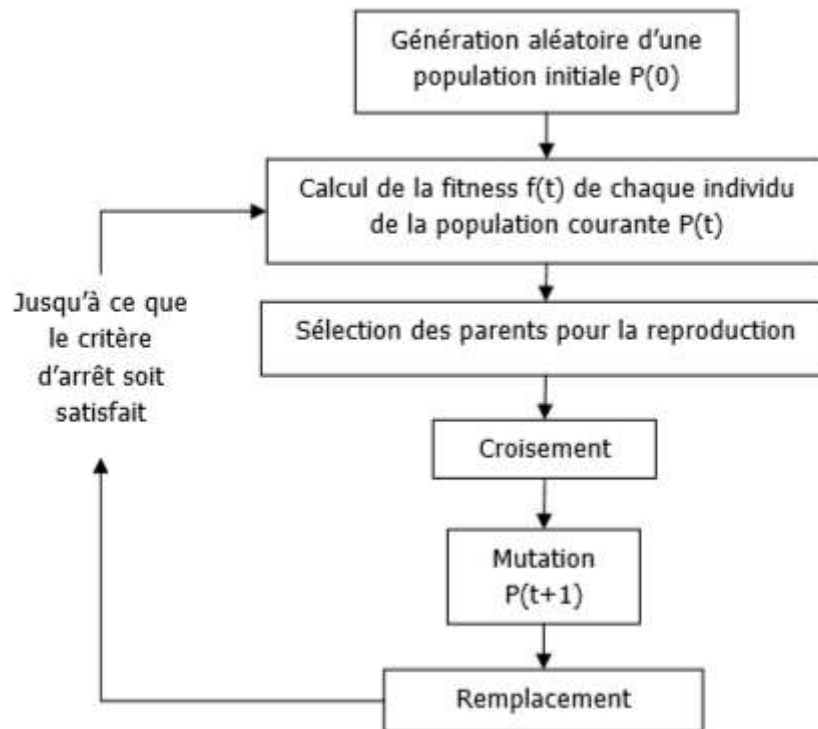


Figure V.5 Fonctionnement général d'un AG de base. Source [57]

- **Opérateurs génétiques**

Les algorithmes génétiques atteignent les meilleurs résultats par l'application des opérations génétiques : opération de sélection, opération de croisement et opération de mutation.

- ✓ **Sélection**

L'opération de sélection sert à sélectionner les parents i représentant des individus élitistes dans la population actuelle, pour permettre générer une progéniture. Nous utilisons un modèle préservé élitiste pour l'opération de sélection et pour cela nous utilisons la corrélation comme mesure d'optimisation et sélection le chromosome (individu optimale). Afin d'obtenir l'individu le plus apte de notre système.

La figure V.6 illustre l'opération de sélection des caractéristiques par un algorithme génétique. Nous appliquons la condition de sélection et boucle jusqu'à l'accès à le critère d'arrêt.

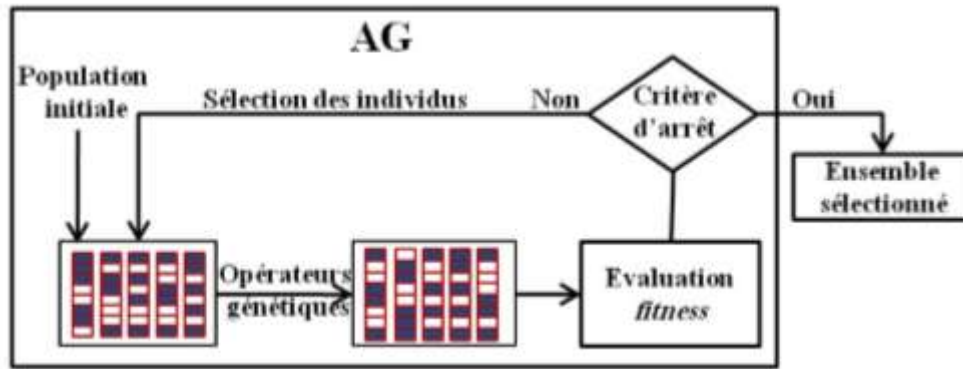


Figure V.6 Sélection de caractéristiques par un algorithme génétique. Source [58]

✓ **Croisement**

Le croisement est chargé de construire un individu (ou une solution) qui soit le mélange de plusieurs solutions. Dans le croisement, les chromosomes échangent des séquences de gènes entre eux. Ce processus est appliqué à chaque paire de chromosomes sélectionnés selon une des méthodes décrites précédemment (corrélation) avec une certaine probabilité P_{crois} . Les paires de chromosomes sont recopiées sans modification dans la génération suivante avec la probabilité. Nous choisissons deux descripteurs de chaque deux image comme parent.

La figure V.7 expose l'opération de croisement entre deux parents.

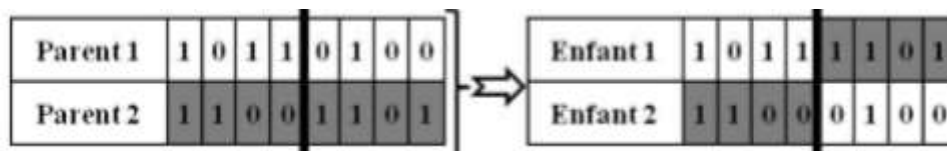


Figure V.7 Opérateur de croisement. Source [58]

✓ **Mutation**

L'opération de mutation est très importante pour conserver les variétés de populations. La mutation est définie comme étant la modification aléatoire de la valeur d'un gène dans un chromosome. La figure V.8 illustre un exemple de mutation appliquée à la quatrième position d'un chromosome binaire. La mutation joue le rôle de bruit qui empêche l'évolution [56].



Figure V.8 Exemple d'une opération de mutation. Source [58]

La figure V.9 illustre le résultat de l'application de l'algorithme génétique. Le premier graphique représente la variation de la fonction fitness en fonction de génération ; dans le deuxième graphique les colonnes en couleurs bleues désignent les neuf descripteurs discriminants obtenus au lieu de 21

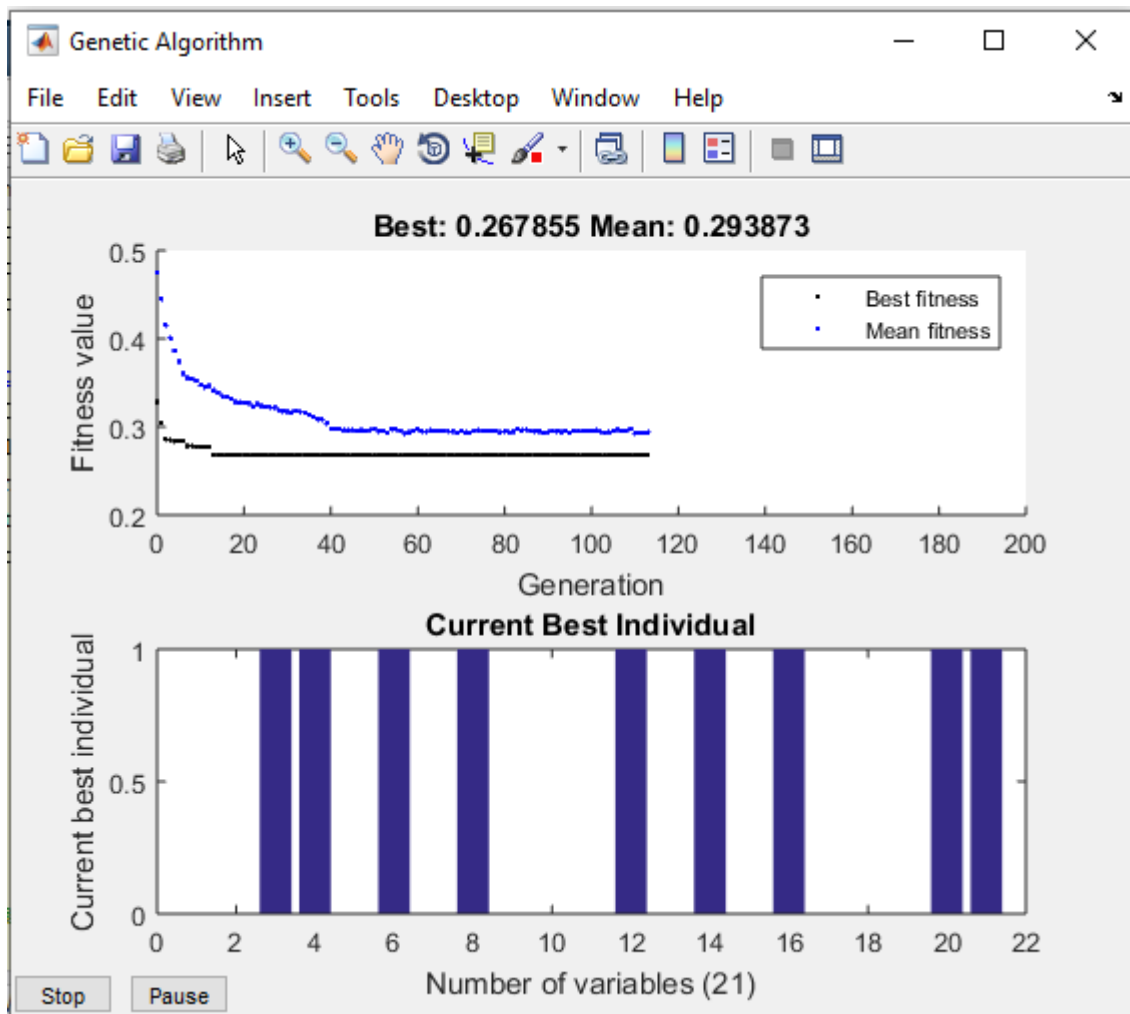


Figure V.9 Résultat d'algorithme génétique

V.5.3 Classification basée Template-matching

Template-matching est une technique de vision par ordinateur de haut niveau utilisée dans la classification des objets et permet d'identifier les parties d'une image (ou de plusieurs images) correspondant au motif d'image donné. Cette technique de modèles compare des parties d'images les unes par rapport aux autres. L'image échantillon peut être utilisée pour

reconnaître des objets similaires dans l'image source. Le but de cette technique est de trouver des occurrences d'un modèle, qui est une sous-image dans une image plus grande. Le processus de correspondance déplace l'image de modèle vers toutes les positions possibles dans une image source plus grande et calcule un index numérique indiquant dans quelle mesure le modèle correspond à l'image située à cette position. La correspondance est faite pixel par pixel [54, 55].

Notre méthodologie de catégorisation des zones d'intérêt segmentées à partir des mammographies est basée la méthode Template-matching, nous utilisons une technique de recherche pour trouver dans la base de données, les zones d'intérêts qui sont pathologiquement similaires à chaque zone d'intérêt requête. Ensuite, nous affectons la zone d'intérêt requête à la classe pathologique la majoritairement renvoyée par le processus de recherche qui sera établie par l'algorithme des K plus proches voisins.

L'algorithme k-plus proche voisin (KPPV) est une méthode de classification supervisée .Elle est basée sur le calcul des distances entre le point et les points de l'ensemble de données d'apprentissage. La mesure de similarité est la distance euclidienne (ou fonction de distance). Ensuite, il assigne le point à la classe parmi ses k plus proches voisins (où k est un entier).

- **Fonction de distance**

Soit A et B représentés par des vecteurs de caractéristiques :

$A = (x_1, \dots, x_m)$ et $B = (y_1, \dots, y_m)$, où m est la dimensionnalité de l'espace caractéristique.

Pour calculer la distance entre A et B, la métrique euclidienne normalisée est généralement utilisée par [59] :

$$dist(A, B) = \sqrt{\frac{\sum_{i=1}^m (x_i - y_i)^2}{m}} \quad (V.11)$$

Les machines à vecteurs de support (SVM) ont été largement utilisées pour la classification supervisée. Elles sont l'une des meilleures techniques de Classification d'images [60].

V.6 Architecture du système Mammo-Zipf-Har

La figure Nous détaillons les deux phases de notre système proposé comme suit :

- **La phase hors ligne**
 - ✓ L'analyse des zones d'intérêt par les lois de Zipf et de Zipf inverse dans le but de tracer les courbes de Zipf et Zipf inverse et extraire les 8 descripteurs texturaux générés par ces lois puissance.
 - ✓ Extraction des 13 descripteurs texturaux générés par la matrice de co-occurrence d'Haralick
 - ✓ Réduction de la dimension du vecteur descripteur réunissant à la fois les descripteurs des lois de Zipf et Zipf inverse ainsi que ceux de la matrice de co-occurrence d'Haralick, à 8 descripteurs suite à l'application des algorithmes génétiques.
 - ✓ Indexation les vecteurs descripteurs dans la base de données.

- **La phase en ligne**
 - ✓ le radiologue introduit une zone d'intérêt requête, segmentée à partir d'une mammographie au système pour déterminer sa pathologie.
 - ✓ L'analyse de la zone d'intérêt requête par les lois de Zipf et de Zipf et l'extraction des descripteurs de texture.
 - ✓ Extraction des descripteurs texturaux relatifs à cette zone d'intérêt à partir de la matrice de co-occurrence d'Haralick.
 - ✓ Application des algorithmes génétiques sur le vecteur descripteur global de la zone d'intérêt requête.
 - ✓ La mesure de similarité (matching) entre le vecteur descripteur de la zone d'intérêt requête et le vecteur descripteur de chaque zone d'intérêt dans la base de données est mesuré par la distance Euclidienne intégrée dans l'algorithme des K plus proches voisin.
 - ✓ Les cas renvoyés par le système suite à la recherche des K plus proches voisin de la région d'intérêt requête s'utilisent comme prédicteurs utiles de la classe pathologique de la requête courante. Ceci, en attribuant à la zone d'intérêt requête la classe pathologique majoritairement renvoyée par le processus de recherche.
 - ✓ Nous avons utilisé aussi le classificateur (SVM) pour le processus de classification afin de rendre notre système fiable et puissant.

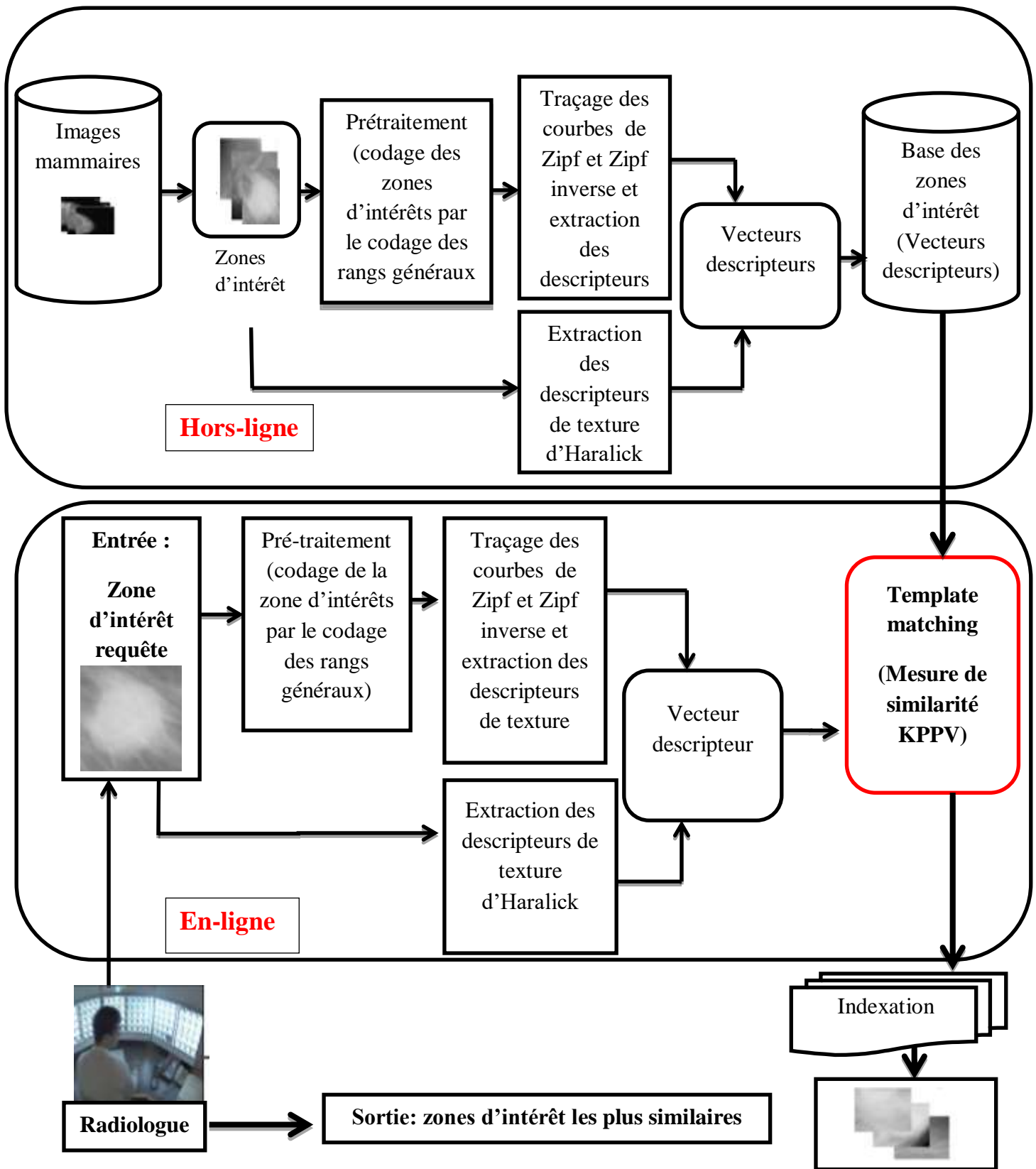


Figure V.10 Architecture du système Mammo-Zipf-Har

V.7 Implémentation

V.7.1 L'environnement de développement

Nous avons implémenté notre système sous un PC performant de type lenovo ThinkPad sous système d'exploitation Windows 10.1 Pro. Processeur est Intel Core i5. Fréquence du processeur : 2,6 GHZ. Mémoire vive : 6 Go et Edition Premium : 64bits.

Le langage de programmation utilisé est : Matlab version 2015a.

MATLAB (MATrix LABoratory) est un environnement puissant, complet et facile à utiliser destiné au calcul scientifique. Il apporte aux ingénieurs, chercheurs et à tout scientifique un système interactif intégrant le calcul numérique et la visualisation. C'est un environnement performant, ouvert et programmable.

On a choisi Matlab parce qu'il possède un approche matricielle permet de traiter les données sans aucune limitation de taille et de réaliser des calculs numériques et symboliques de façon fiable et rapide. Grâce aux fonctions graphiques de MATLAB, il devient très facile de modifier interactivement les différents paramètres des graphiques pour les adapter selon nos souhaits. Dans MATLAB, aucune déclaration n'est à effectuer sur les nombres. En effet, il n'existe pas de distinction entre les nombres entiers, les nombres réels, les nombres complexes et la simple ou double précision. Cette caractéristique rend le mode de programmation souple et très rapide. En Fortran par exemple, une subroutine est presque nécessaire pour chaque variable simple ou double précision, entière, réelle ou complexe. MATLAB a aussi une grande variété de fonctions utiles pour le praticien en algorithmes génétiques et pour ceux qui souhaitent expérimenter avec l'algorithme génétique pour la première fois.

Génétic Algorithm Toolbox est un ensemble de routines, écrites principalement en fichiers m-fichiers, qui implémentent les fonctions les plus importantes des algorithmes génétiques.

V.7.2 La base des images mammaires MIAS

Pour élaborer notre travail, nous utilisons 120 images mammaires qui sont déjà diagnostiquées et archivées dans la base MIAS (Mammograms Image Analysis Society).

Les images mammaires sont divisées en pathologie connues 59 images mammaire sans lésions, 37 images mammaires portant une tumeur bénignes et 24 portant une tumeur maligne. Notons que la méthode de segmentation des zones d'intérêt a été présentée dans [5].

V.7.3 Evaluation des performances

L'évaluation de notre système est réalisée par la validation croisée. Cette évaluation est un moyen de prédire l'efficacité d'un modèle dans la pratique. Le principe de cette validation consiste à diviser la base de données MIAS en deux sous échantillons.

Le premier dit d'apprentissage (communément supérieur à 60 % de l'échantillon) et le second dit de test.

Le critère de performance du système proposé a été évalué par la précision de la classification (%), voir plus, le temps requis au système pour la classification d'une zone d'intérêt.

Nous illustrons sur le tableau V.2 la variation de la précision de classification du système «Mammo-Zipf-Har» pour diverses valeurs des K plus proche voisins et sans la réduction des vecteurs descripteurs par les algorithmes génétiques.

Tableau V.2 Variation de la précision de classification du système «Mammo-Zipf-Har» sans la réduction des vecteurs descripteurs par les algorithmes génétiques.

Les valeurs de K	Précision de la classification (K-PPV)	Temps d'exécution	Précision de la classification (SVM)	Temps d'exécution
2	70,21%	0,74	57%	2,75
3	61,70%	0,82		
4	63,82%	0,88		
5	55,31%	0,91		
6	53,19%	1,04		

Tandis que sur le tableau V.3, nous exposons la précision de classification du système «Mammo-Zipf-Har» pour diverses valeurs des K plus proche voisins et suite à la réduction des vecteurs descripteurs par les algorithmes génétiques.

Les valeurs du K	Précision de la classification (K-PPV)	Temps d'exécution	Précision de la classification (SVM)	Temps d'exécution
2	80,85%	0,49	61,70%	1,20
3	72,34%	0,67		
4	70,21%	0,74		
5	72,34%	0,76		
6	72,34%	0,87		

Tableau V.3 Variation de la précision de classification du système «Mammo-Zipf-Har» après avoir réduit les vecteurs descripteurs par les algorithmes génétiques.

Nous pouvons affirmer que la précision de la classification varie entre 70.21% et 80.85% et que la meilleure performance du système proposé a été obtenue pour k=2.

Les résultats montrent que la performance du système «Mammo-Zipf-Har» est meilleure lorsque nous appliquons les algorithmes génétiques pour l'obtention des descripteurs discriminants.

Nous avons comparé nos résultats obtenus avec d'autres approches comme indiqué sur le tableau V.4 suivant :

Approche	Base de données	Précision de classifications (%)	
		SVM	K-PPV
Zipf, Zipf inverse et Haralick	MIAS	61.70%	80.85%
Zipf, Zipf inverse et LBP	MIAS	76%	69%
Zipf, Zipf inverse et les SFTA	MIAS	60%	62%
Zipf, Zipf inverse et les filtres de Gabor	MIAS	67,5%	65%
Zipf, Zipf inverse et les Ondelettes de Haar	MIAS	70%	80%
Zipf, Zipf inverse et la transformée discrète de Fourier	MIAS	62%	80%

Tableau V.4 L'approche proposée vis-à-vis d'autres approches.

V.8 Présentation de quelques vues du système «Mammo-Zipf-Har»

La première fenêtre de notre système «Mammo-Zipf-Har» appelée login, nous permet d'accéder aux différentes fonctionnalités du système. Ceci est mentionné sur la figure V.10.



Figure V.11 Fenêtre Login du système «Mammo-Zipf-Har»

En cliquant sur le bouton «Entrer», nous accédons à l'interface principale du système «Mammo-Zipf-Har», après avoir introduit le nom d'utilisateur et le mot de passe comme mentionné sur la figure V.11.



Figure V.12 Interface principale du système «Mammo-Zipf-Har»

En cliquant sur le bouton « Parcourir », le radiologue choisira la zone d'intérêt requête qu'il désire diagnostiquer.

Sur figure V.12, nous démontrons également le déroulement du processus de classification des images mammaires requêtes dans le système Mammo-Zipf-Har proposé. En effet, le radiologue introduit la zone d'intérêt requête au système et cherche les régions d'intérêt similaires afin de récupérer le résultat de la classification par les deux algorithmes des K-PPV ainsi que les SVM : tumeur maligne ou bien bénigne ou tissu de parenchyme normal.

Comme mentionné sur la figure V.12 où 2 régions d'intérêt présentant un tissu normal ont été renvoyées par le processus de recherche, aidant le radiologue à diagnostiquer la région d'intérêt requête comme tissu parenchyme sein ce qui est vraiment le cas car nous pouvons constater que la zone d'intérêt requête appartient à la classe ne présentant pas de pathologies

Chapitre V : Conception et Implémentation

mammaires. En outre, le système Mammo-Zipf-Har offre un affichage des voisins appartenant à la même classe pathologique de la région d'intérêt requête.

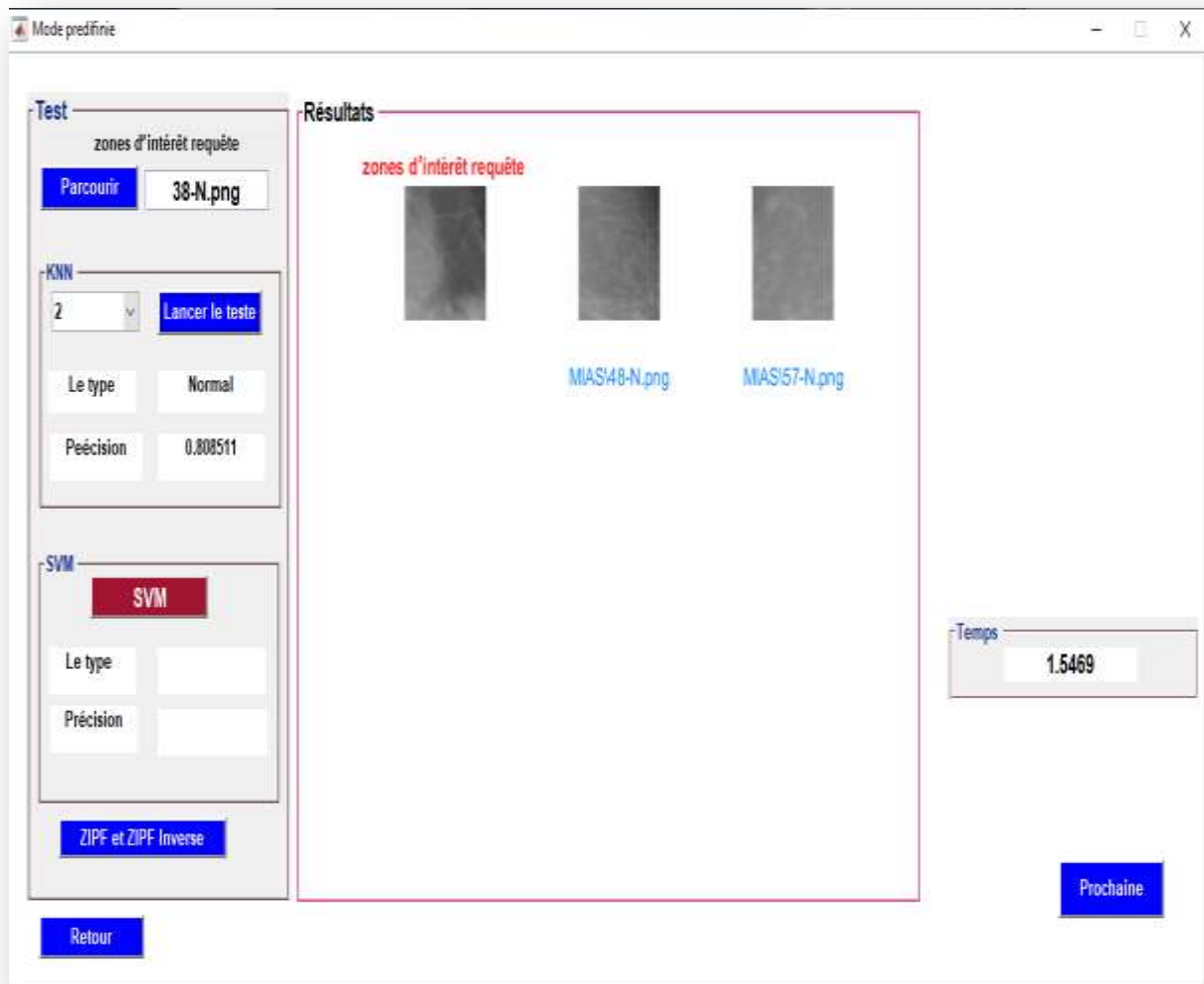


Figure IV.13 Processus de décision médicale issue de l'utilisation de la sortie de Mammo-Zipf-Har par la considération de $k=2$.

Sur la figure IV.14, nous démontrons également le déroulement du processus de classification des images mammaires requêtes dans le système Mammo-Zipf-Har proposé par la considération de $K=5$.



Figure IV.14 Processus de décision médicale issue de l'utilisation de la sortie de Mammo-Zipf-Har par la considération de $k=5$.

V.9 Conclusion

Dans ce chapitre, nous avons détaillé les phases d'élaboration de notre système d'indexation et de recherche des mammographies par le contenu «Mammo-Zip-Har» qui nous a permis de classifier les images mammaires en utilisant la technique Template-matching à travers le classifieur des K plus proches voisins (K-PPV) ainsi que celui des machines à vecteur de support (SVM). Nous avons obtenu une précision de classification assez bonne de l'ordre de 80.85%, ceci par les K-PPV et de l'ordre de 61.70 % pour les SVM.

Conclusion générale

CONCLUSION GÉNÉRALE

Vu le nombre important d'images médicales réalisées par jour, ceci, a rendu l'exploitation efficace de ces ressources un défi pour les experts du domaine. En effet, le besoin de la conception des systèmes bien définis qui permettent un stockage et une recherche efficaces de ces images était en ordre.

L'enjeu principal de la vision par ordinateur est d'extraire les propriétés visuelles des tissus consistants en de descripteurs numériques pertinents et distinctifs qui permettent la reconnaissance d'objets ou la recherche d'images par contenu. En effet, l'image médicale ne peut effectivement s'appréhender qu'au moyen de la chaîne de production et d'interprétation des images.

La présence des lésions qui sont obscurcies et spécifiques ou bien possédant des apparences identiques sur les mammographies est un obstacle vis-à-vis la suffisance des descripteurs géométriques pour aboutir à une distinction précise entre les masses bénignes et malignes d'où la puissance des descripteurs de la texture. Effectivement, l'analyse de la texture est très importante pour la vision physique humaine, ceci dit, les descripteurs de l'analyse de la texture ne sont pas facilement perçus et quantifiés par la vision humaine.

Dans ce mémoire, nous avons abordé le problème de l'aide au diagnostic médical du cancer du sein assisté par ordinateur. Pour cela, nous nous sommes basé sur l'indexation et la recherche d'images par contenu, tout en appliquant simultanément deux techniques d'analyse de la texture des régions d'intérêt obtenues suite à une segmentation des mammographies, consistant en les lois puissance Zipf et Zipf inverse qui ont prouvées récemment leur puissance en analyse d'image ainsi que la matrice de co-occurrence d'Haralick étant riche en information de texture et devenant la plus connue et la plus utilisée pour extraire des descripteurs de texture. En outre, une phase de sélection des descripteurs les plus discriminants a été réalisée par les algorithmes génétiques.

Le système développé « Mammo-Zipf-Har » d'indexation et de recherche des mammographies par le contenu se base sur l'approche Template matching qui affecte la région d'intérêt en cours d'analyse (requête) à la classe pathologique la majoritairement renvoyée par le processus de

Conclusion générale

recherche, où nous avons atteint une précision de classification assez bonne de l'ordre de 80,85% par l'algorithme des K plus proches voisins(KNN), et une précision de classification de l'ordre de 61,70% par l'algorithme des machines à vecteur de support (SVM).

Sans doute l'indexation et la recherche des mammographies par le contenu demeure la plus adéquate aux radiologues vu qu'elle reflète les standards adoptés par ces derniers pour classifier les lésions mammaires en bénignes ou malignes, ceci au biais du diagnostic à base de cas.

Selon les radiologistes, l'aide au diagnostic médical assisté par ordinateur (CAD) est d'une grande utilité pour le soutien des contrôles médicaux préventifs en mammographie. En outre, divers grands domaines comme la neurologie ou le cardio-vasculaire suivront à leur tour l'innovation technologique fournissant au praticien des nouvelles approches de diagnostic.

Perspectives

Dans nos futurs travaux, nous étudierons de nouvelles techniques de caractérisation de la texture, de sélection des descripteurs les plus discriminants ainsi que de classification.

BIBLIOGRAPHIE

- [1] Christine Graffigne, Annick Montanvert, thèse HAL Id: hal-00706168 : Analyse d'images: Filtrage et segmentation, 23 Jun 2017
- [2] Ramla Intissar et ZAIR Asmae Mama, Mémoire de Master : Detection des lésions de l'abdomen et reconstruction tridimensionnelle, 22 Mai 2016.
- [3] Bendaoud Mohammed Habib, thèse : Développement de méthodes d'extraction de contours sur des images à niveaux de gris , 27/02/2017
- [4] Laurent Guigues ; thèse de l'Université de Cergy-Pontoise : Modèles Multi-Échelles pour la Segmentation d'Images ; Décembre 2003
- [5] Meriem Hamoud, Indexation et segmentation d'images basées loi de Zipf et Zipf inverse, mémoire de doctorat ism2015
- [6] Badreddine Belkis ,DoudI Hayet , Analyse de la texture des images mammaires par une fusion des lois de Zipf et des SFTA pour la classification des tumeurs mammaires via l'analyse en composantes principales ,mémoire de master 2018
- [7] Kidiyo Kpalma, thèse HAL Id: tel-00446030 : Analyse et Interprétation d'Images: Analyse Multiéchelle pour la Description des Formes Planaires, 11 Jan 2010
- [8] TERKI Hakima et Benyelles Khadidja, Mémoire master : Analyse des images mammographiques en vue de la détection et la caractérisation des Microcalcifications mammaires, 23 juin 2013.
- [9] Antoine Lefebvre, Contribution de la texture pour l'analyse d'images à très haute résolution spatiale : application à la détection de changement en milieu périurbain, thèse doctorat Université Rennes 27 May 2011.
- [10] hariz bekkar ouacila , Belbey Medjda, Analyse de la texture des images mammaires par une fusion des lois de Zipf et des Ondelettes pour la classification des tumeurs mammaires via l'analyse en composantes principal, 2018.

- [11] S. Dash. M.R. Senapati. U.R. JenaK-NN based automated reasoning using bilateral filter based texture descriptor for computing texture classification. Egyptian Informatics Journal, (2018).
- [12] Danuser, G. Computer Vision in Cell Biology. Elsevier Inc. (2011).
- [13] Brian L. DeCost, Elizabeth A. Holm. A computer vision approach for automated analysis and classification of microstructural image data. Computational Materials Science, (2015)
- [14] Isabelle Claude, Analyse de textures en traitement d'images journal GB06
- [15] Sevcan Aytac Korkmaz Mehmet Fatih Korkmaz, Journal: A New Method Based Cancer Detection in Mammogram Textures by finding Feature Weights and using Kullback-Leibler Measure with Kernel Estimation, 2015.
- [16] Aditya A Shastri Deepti Tamrakar, Kapil Ahuj: Density-wise two stage mammogram classification using texture exploiting descriptor, , Journal , january 2018.
- [17] Arnaud Boucher, Recalage et analyse d'un couple d'images : application aux mammographies, thèse doctorat Université Paris Descartes, 9 janvier 2013
- [18] Pol Kennel, Analyse de texture pour la classification et segmentation d'images, Séminaire "Imagerie pour les Plantes et les Paysages" (I2P) de l'AMAP, 24 Juin 2013.
- [19] Hanifi Majdoulayne, Extraction de caractéristiques de texture pour la classification d'images satellites, thèse de doctorat de l'université de toulouse, 02 / 11 / 2009.
- [20] Aimilia Gastounioti, Emily F. Conant and Despina Kontos, journal : Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment, 2016.
- [21] Santa Di Cataldo, Elisa Ficarra, Mining textural knowledge in biological images: Applications, methods and trends, journal 2017
- [22] Ali Ismail, Journal : Image Feature Detectors and Descriptors Foundations and Applications, November 2015.
- [23] Dayakshini Sathish, · Surekha Kamath , Journal : Asymmetry analysis of breast thermograms using automated segmentation and texture features, 2017.

- [24] Rodrigo Pereira Ramos, Marcelo Zanchetta do Nascimento, Danilo Cesar Pereira , Texture extraction: An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms, journal homepage: www.elsevier.com/locate/eswa .2017
- [25] Rodrigo Pereira Ramos , Journal : Texture extraction : An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms, 2012.
- [26] Geert Litjens, Thijs Kooi, Babak Ehteshami, Joinal : A survey on deep learning in medical image analysis, 26 July2017.
- [27] Maroua Mehri, Mohamed Mhiri, Petra Gomez-Krämer, thèse : Étude comparative de trois ensembles de descripteurs de texture pour la segmentation de documents anciens, 24 Mar 2014.
- [28] Edson Damasceno Carvalho, Antonio Oseas de Carvalho Filho, Journal : Methodofdifferentiation ofbenign andmalignant masses indigital mammograms using texture analysis based on phylogenetic diversity, 2018.
- [29]Ludovic Paulhac, thèse :Outils et méthodes d'analyse d'images 3D texturées: application à la segmentation des images échographiques, HAL Id: tel-00576507, UNIVERSITÉ françois rabelais de tours, 14 Mar 2011.
- [30] Yves CARON, Contribution de la loi de Zipf à l'analyse d'images, thèse doctorat, L'UNIVERSITE DE Tour 2004.
- [31] Elisa Ficarra, Mining textural knowledge in biological images: Applications, methods and trends, November 2016.
- [32] Alexandra Schaffar, La loi de Zipf dans la science régionale : entre anciennes controverses et nouvelles perspectives, LIM / Université de La Réunion LEAD / Université de Toulon2007
- [33] Laimeche Lakhdar ,détection des information cachées dans les images numérique basée sur la loi de Z,ipf ,thèse université cheikh Larbi Tbessi, Tebessa.2010
- [34] Honor G.K , Zipf law , Glottometrics , journal scientifique,2002
- [35] Alexandra Schaffar, Les Lois De Zipf Et De Gibrat Pour Les Villes : UNE Introduction, Région et Développement n° 36-2012
- [36]Hamoud,M. Merouani, H. F. Detection of a Region of Interest in the Images Based on Zipf Laws. In The Seventh IEEE. (SITIS'2011). Dijon, France (2011).

- [37] Benmazou Sarah, thèse : Indexation des images basees automates cellulaires et agents ; université de Badji Mokhtar Annaba , 2015
- [38] Remontet L, Esteve J, Bouvier AM, et al. Cancer incidence and mortality in France over the period 1978-2000. Rev Epidemiol Sante Publique 2003.
- [39] Merabet Chahrazed, Mémoire final : Analyse de la texture des images mammaires par une fusion Des lois de Zipf et de la transformée de Fourier pour la classification des tumeurs mammaires via L'analyse en composantes principales, université d'El-oued, 2018.
- [40] Imene cheikhrouhou epse kachouri, Description et classification des masses mammaires pour le diagnostic du cancer du sein , universite d'evry val d'essonne,2012.
- [41]Médecin généraliste, Journal : Cancers du sein /Du diagnostic au suivi, Institut national de cancer, France, 2014.
- [42] Médecin généraliste, Journal : Dépistage des cancers du sein s'informer et décider, Institut National de cancer du sein France, 2015.
- [43] Chabane siham, Haddache Narimane, Mémoire de master : Segmentation d'images mammographies, Université Abderrahmane Mira Béjaïa, 2017.
- [44] Bennai Mohamed Taher, Réalisation d'un système multi-agents adaptatifs pour l'imagerie médicale, université M'hamed Bougara-Boumerdes, 2013.
- [45] Hiroshi Fujita , Takeshi Hara, Journal : Breast Imaging, Gifu City, Japan, July 2, 2014 .
- [46]Oliver Faust, U.Rajenda, Journal : Comparative assesment of texture features for the identification of cancer in ultrasound images : a review, the Polish Academy of Sciences, 2018.
- [47] Edson Damasceno Carvalho, Journal : Method of differentiation of benign and malignant masses in digital mammograms using texture analysis based on phylogenetic diversity, University of Piauí, Rua Cicero Duarte, SN, Campus de Picos, Junco, 2018.
- [48] Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley , A Deep Learning Approach for the Analysis of Masses in Mammograms with Minimal User Intervention, American,2016.
- [49] S.M.A Behechti, Journal : Classification of abnorm alities in mammograms by new asymmetric fractal features, Univercity of Medical Iran, 2015.
- [50] Janine Katzen , journal homepage: www.elsevier.com/locate/clinimag : A review of computer aided detection in mammography, United States of America, 2018.

- [51] Johanna O. P. Wanders, Carla H. van Gils, journal : The combined effect of mammographic texture and density on breast cancer risk: a cohort study ,2018.
- [52] S.W. Zucker and D. Terzopoulos. Finding structure cooccurrence matrices for texture analysis. Computer Vision Graphics and Image Processing ; 2000.
- [53] Aditya A Shastri Deepti Tamrakar, Kapil Ahuj, Journal : Density-wise two stage mammogram classification using texture exploiting descriptor, january 2018.
- [54] Nazil Perveen, Darshan Kumar and Ishan Bhardwaj, , international journal : An Overview on Template Matching Methodologies and its Applications, Raipur India, 2013.
- [55] Nazanin Sadat Hashemia, Roya Babaei Aghdamb, Atieh Sadat Bayat Ghiasi, American Scientific Research Journal : Template Matching Advances and Applications in Image Analysis, 2013.
- [56] MENGHOUR Kamilia , thèse : Approches Bio - inspirées pour la Sélection d'Attributs, université de Badji Mokhtar Annaba ,2015
- [57] 2015Jianjiang lu, Tianzhon Zhoa, journal : featur selection based on genetic algorithm for image annotation, 2008.
- [58] Hassan CHOUAIB, thèse doctorat : Sélection de caractéristiques: méthodes et applications, université Paris, 8 juillet 2011.
- [59] Li-Yu, H. Min-Wei, H. Shih-Wen, K. Chih-Fong, T. The distance function effect on k-nearest neighbor classification for medical datasets.. (2016).
- [60] Bahjat Safadi, thèse : Indexation sémantique des images et des vidéos par apprentissage actif, HAL Id: tel-00766904, décembre 2012.
- [61] Y. Guermeur and A. Elisseeff and H. Paugam-Moisy. A new multi-class SVM based on a uniform onvergence result. In International Joint Conference on Neural Networks, volume IV (2000).
- [62] J. Weston and C. Watkins. Multiclass support vector machines. In proceedings of ESANN99, Brussels. D. Facto Press, 1999.
- [63] Marcelo Zancheta ,Alessandro Santana , Journal : Classification of masse in mammographic image using wavelet domain features and polynomial classifier.2015